DOCUMENT RESUME

ED 195 952 CS 005 764

AUTHOR Palmer, John C.; And Others
TITLE Is Visual Information Processing Related to Reading?
INSTITUTION Washington Univ., Seattle.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE [79]
GRANT NIE-G-77-0012
NOTE 132p.

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS Auditory Stimuli; *Cognitive Processes; College Students; Higher Education; Listening Comprehension; *Reading Comprehension; *Reading Processes; *Reading Research; Reading Skills; *Visual Perception; Visual Stimuli

ABSTRACT

A large stratified sample of university undergraduate students differing in reading ability performed a diverse set of psychometric and information processing tasks in a study exploring the role of visual information processing skill as a component in reading ability. Using a correlation analysis of individual differences, the interrelationships among the various measures were examined, treating reading time and reading comprehension separately. Two central findings emerged. First, reading comprehension shares the same processes as listening comprehension, and this cannot be accounted for by general intelligence. Second, visual information processing predicts reading ability considerably better when the units of analysis are words as opposed to letters; this prediction improves with the size and complexity of the lexical unit. A multivariate view of reading best accommodates the findings. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED195952

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Is Visual Information Processing Related to Reading?

John C. Palmer

University of Michigan

Colin M. MacLeod

Scarborough College, University of Toronto

Earl Hunt and Janet E. Davidson

University of Washington

Short Title: Visual Processing and Reading

Send Proofs to: John C. Palmer

Human Performance Center

330 Packard Road

University of Michigan

Ann Arbor, MI 48104

U. S. A.

CS005764

Abstract

A large stratified sample of university undergraduates differing in reading ability performed a diverse set of psychometric and information processing tasks. The goal was to explore the role of visual information processing skill as a component in reading ability. Using a correlation analysis of individual differences, the interrelationships among the various measures were examined, treating reading time and reading comprehension separately. Two central findings emerged. First, reading comprehension shares the same processes as listening comprehension, and this cannot be accounted for by general intelligence. Second, visual information processing predicts reading ability considerably better when the units of analysis are words as opposed to letters; this prediction improves with the size and complexity of the lexical unit. A multivariate view of reading best accommodates the findings.

Reading, because of its clear social importance, poses a unique challenge to cognitive psychologists. Can an analytic, information-processing approach be applied successfully to further our understanding of such a complex activity? If so, what are the component processes involved in reading? How do good and poor readers differ in their assembling of these processes? Although we cannot expect to respond to such questions with any certainty, we can begin to provide some preliminary answers and to eliminate some potential mechanisms from further consideration. At a very modest level, that is what we hope to accomplish here.

Consider first the issue of whether there is any relationship between reading and performance on information-processing tasks. Apparently, it is a widely held belief that there is a relationship. In accord with this belief, we will discuss below the recent work of Jackson and McClelland (in press), Frederiksen (Notes 1, 2) and Perfetti and his coworkers (Perfetti & Lesgold, 1977; Perfetti, Bell, Hogaboam & Goldman, Note 3). However, these represent virtually the entire set of relevant studies. Even considering other isolated reports of correlations between reading ability and a single cognitive task (e.g., sentence-picture verification in MacLeod, Hunt & Mathews, 1978), the supporting evidence is surprisingly sparse. Thus, one aim of the present study

is to help strengthen this empirical base.

To guide our work, we have formulated a single question -- "Is visual information processing related to reading?" Perhaps the best way to begin is by unpacking the component terms in this question. The easiest term to define in our question is "related", which we will do operationally: A relationship exists between reading ability and some cognitive skill when the two are significantly correlated over a group of individuals carefully selected to represent a normal sample of readers. Unfortunately, the other terms in the question defy such simple definition. An outline of our overall approach is necessary before tackling "reading" and "visual information processing".

Consistent with earlier work (cf. Hunt, 1978), we have adopted an individual differences approach, taking as our basic data the correlations between individual reading ability and performance in selected information processing tasks. This is in sharp contrast to the nomothetic experimental approach, which can also be used in answering "shared process" questions.[1] As Underwood (1975) has pointed out, if two tasks share the same underlying mechanisms, then the individual variation in performance on those two tasks should be correlated. Consequently, individual differences are an important aspect of theory construction. However, problems do arise

in that mechanisms other than those of immediate interest may also produce significant correlations. Part of the ingenuity in such research must lie in employing appropriate theoretical tools and analyses to disentangle the mechanisms.

The isolation of the mechanism of interest is based upon three principal considerations. First, the contribution of random error is estimated from parallel scores on each task. Second the contribution of systematic error from uninteresting response components is estimated from covariate measures of response performance (e.g., choice reaction time). Finally, the generality of the underlying trait(s) is estimated from a set of proposed converging measures. These considerations are combined in an analysis that estimates the "true score" representing a mechanism's variation in performance. If the model underlying this analysis is correct, then Underwood's strategy is viable.

READING

There are two other terms in our question left to define, "reading" and "visual information processing". We will begin with reading. Gibson and Levin (1975) say that "reading is extracting information from text . . . an active process, self-directed by the reader in many ways and for many purposes (p. 5)." This appears straightforward; however, this definition is more

problematic if one considers what reading is not. There are really two subproblems for us in asking what reading is not: (1) Can an individual's reading ability be captured with a single score? and (2) To what extent is reading simply a reflection of more general intellectual capacities? We will discuss both of these problems in some detail, since they play central roles in our research.

We frequently talk about good and poor readers as if reading ability were unidimensional. Immediately, though, the distinction between reading speed and reading comprehension intrudes, and we recognize that reading is not a single, unified package. Unfortunately, much of the literature relies on reading tests that require either a comprehension component in the speed measure or a time limit on the comprehension measure (cf. Tinker, 1939). The similarity of prediction for speed and comprehension is, as a result, spurious; it is based on a confounding. When this confounding is avoided, the correlation between speed and comprehension is typically very low (e.g., Blommers & Lindquist, 1944). In fact, as is widely observed in studies of simpler cognitive tasks, speed can be traded for accuracy (and vice versa), resulting in a negative correlation between speed and comprehension (e.g., McConkie, Rayner & Wilson, 1973). It may even be that the ability to adjust reading speed to the comprehension demands of the situation is a stable

individual trait (cf. Gibson & Levin, 1975). Because of this complicated relationship between speed and comprehension in reading, studies of individual differences must include measures of both variables. Reading ability cannot be captured with a single score.

We have been speaking of "reading skill" as a trait that is apart from any other individual capacities. In educational psychology, considerable stress is placed upon the distinction between reading skill and various concepts of intelligence. "Intelligence", however, is itself a rather vague catch-word used to refer to performance on intellectual tasks, and it is well known that virtually any two intellectual measures of performance are likely to have moderate positive correlations. The situation is even murkier if we wish to make a distinction between reading ability and verbal intelligence. The latter concept refers to one's ability to deal with problems that are presented using language stimuli, and reading is certainly an integral part of any such task. There is a conceptual lack of clarity in thinking about reading skill as being completely apart from verbal intelligence. One can, however, think of studying individual differences in those aspects of the reading task that require skills not usually associated with the concept of general (or verbal) intelligence. Thus, a conceptual distinction can be made between measures of intelligence

and measures of some components of the total process that we call reading. If we are concerned with the reading process as a whole, though, we must also be concerned with general intellectual competence.

VISUAL INFORMATION PROCESSING

An information processing model of reading, regardless of its specific form, makes the assumption that reading can be understood by analyzing it into its component processes. Typically, this consists of a set of serially executed stages (e.g., Gough, 1972; Mackworth, 1972). This represents an extreme analytic position -- reading takes place in a sequence of independent processing steps, beginning with the detection of visual stimuli, proceeding through pattern recognition and memory manipulation, and eventually reaching the level of abstracting meaning. Despite the fact that this extreme analytic view is almost certainly wrong, the basic information processing approach is potentially very useful in studying reading. Its greatest advantage is in breaking the huge general problem of reading into manageable units.

We have chosen to single out visual information processing. By this term, we mean the decoding and representation of orthographic units, and the subsequent manipulation of those simple representations. In other words, our interest centers on pre-comprehension processing of orthographic strings, the initial stages of

reading. As was the case in defining reading, there are really several sub-problems in isolating visual information processing. In passing, we will point to those that have motivated a number of the measures that we collected and that dictated several of our analysis techniques.

There is a large body of literature in experimental psychology that examines letter and word recognition and related processes (for reviews, see Adams, Note 4; Smith & Spoehr, 1974). Clearly, these nomothetic studies are aimed at just the processes we are interested in, those that should be critical in the initial stages of reading. For this reason, we will briefly overview the principal characteristics of these studies.

First, consider the types of questions that are being asked. There is widespread concern over the units -- letter features, letters, letter clusters, or words -- involved in perception and recognition. At the same time, of course, the nature of the processes used in extracting these units is of central importance. Are they serial or parallel, interactive or independent? What role do expectancies (e.g., syntactic, semantic) play in perception? Smith and Spoehr (1974) point out these and other questions and illustrate the range of theories that have been offered as possible answers. Our concern with this work takes a very pragmatic form; we wished to choose tasks for our study of individual differences which seemed

representative, and for which there were existing models. This resulted in a set of very standard paradigms being chosen, including search, matching, and span tasks.

To examine visual information processing in isolation would be virtually impossible. Instead, we attempted to arrange individual experiments within our study in such a way as to permit estimates of its contribution to be derived. Of course, this is particularly crucial where the individual's data are of paramount concern. We took several steps to minimize other influences. First, a two-choice reaction time measure was collected to be used as a covariate; because most of our tasks involved reaction time measurement, it was essential to factor out an individual's general facility in dealing with speeded tasks. This is a rather straightforward precaution.

Second, we varied the size of the stimulus unit, examining letters and words over a constant set of tasks. These tasks were chosen especially to vary in the extent to which they relied on purely visual information processing. Our fundamental notion was that, as the size and the complexity of the unit increased, the pattern of correlations of task performance with reading ability should change. If visual information processing is an especially crucial component of reading, the correlations should decrease with unit size. If visual information processing is only one of several key processes, the correlational

pattern could well not change with unit size. Finally, if visual information processing is relatively unimportant in reading (but other more complex processes are important), the correlations should increase with unit size. This factorial combination of different tasks and units is a major innovation in the present study.

A third design feature of the study was the administration of parallel comprehension tests, one based on reading and one based on listening. This was done to ascertain whether the comprehension processes in reading were general or unique. [The logic here is similar to that invoked earlier for the utility of obtaining an intelligence-test score]. With the listening test partialled out, we could isolate whatever components of comprehension remained as specific to reading. This would also help by placing an upper limit on the unique contribution of visual information processing to reading. Our concern was the observation that reading difficulty might reflect reduced ability to comprehend language, not to decode written symbols (i.e., not visual information processing). The relationship between listening and reading comprehension is frequently discussed (e.g., the papers in Kavanagh & Mattingly, 1972; Sticht, 1972), but only infrequently directly examined. Furthermore, results do conflict, with some investigators arguing that there are two types of comprehension (e.g., Guthrie

& Tyler, 1976), and others arguing that there is only one type (e.g., Smiley, Oakley, Worthen, Campione & Brown, Note 5). An empirical test seems necessary if we are to distinguish visual information processes and comprehension processes in reading.

To this point, we have said relatively little about prior research on visual information processing and reading. Although there are a number of nomothetic studies (e.g., Estes, 1975), there have been few individual differences studies. Generally, these have been of the single experiment variety, using an extreme groups design (e.g., Mason & Katz, 1976). While these are useful in their own domain they are not able to address the broader issues with which we are concerned. However, three recent larger-scale correlational projects provide precedents for our work, and warrant extended discussion here.

The most complete of the current efforts is that of Jackson and McClelland (in press; see also 1975). They set out to find the best predictors of adult "effective reading speed" (percent correct on a comprehension test times the speed measure) among a set of a dozen cognitive measures. From our perspective, the most interesting twist to their study was the use of a listening comprehension measure to determine the relative contribution of general verbal abilities. Over their full sample of 52 university-student readers, they found strong relation-

ships between each of their two reading measures and one listening measure (r = .67, .59). They then selected two 12-subject extreme groups and demonstrated that several letter and word recognition tasks were moderately correlated with effective reading speed. Of these, the letter-name matching task (cf. Posner, Boies, Eichelman & Taylor, 1969), combined with listening comprehension, appeared to form the best predictor for effective reading speed. Jackson and McClelland went on to propose that their general comprehension and letter-code components are causally related to reading speed.

Four problems, all acknowledged by the authors, affect the evaluation of the Jackson and McClelland study. The first two are related to the sample. One must certainly be cautious in interpreting any correlation based on a small sample of only 24 subjects. Compounding the size of sample problem is its composition. Their extreme groups design undoubtedly inflated their correlations, although we cannot determine the extent. For example, the relationship between reading and listening is available for both samples. It was accentuated from r = .67, .59 in the full sample to r = .75, .70 in the extreme groups only. The second two problems are related to the data collected. In any study of individual differences, a major concern must always be the reliability of the individual measures to be correlated. Consider the

"effective reading speed" composite in Jackson and McClelland's study. When the two components are examined separately, reading speed (r = .82) is found to be much more reliable than reading comprehension (r = .55) over the whole sample. Thus, the more reliable speed measure dominates in the composite, making it risky to generalize at all to comprehension. A closely related problem is the validity of the measures. Converging tasks are essential in identifying the most appropriate underlying mechanism as a predictor; the task-mechanism distinction is impossible without such convergence. Unfortunately, the letter-name matching task was unique; we cannot identify the trait underlying its contribution to the prediction of reading. Each of these problems places a restriction on the domain and strength of Jackson and McClelland's conclusions.

Two further reports, although not as directly related, are also of note. Frederiksen (Notes 1, 2) has conducted a factor-analytic study using a battery of information-processing and reading tasks. On the basis of this, he has developed a five-element model of the adult reading process, testing alternative models using Joreskog's (1970) methods. While his model provided a fair fit of the data, the conclusions are again weakened by a very small sample size (20) and lack of adequate reliability in five of the eleven measures

taken. Despite these cautionary notes, the specificity of the model and potential power of the analysis makes this a very interesting study.

The remaining approach stems from the work of Perfetti and his associates which contrasts groups of young readers (Perfetti & Hogaboam, 1975; Perfetti & Lesgold, 1977; Perfetti et al., Note 3). They have found consistent reading level differences in a set of letter and word tasks. Like Jackson and McClelland, however, they have failed to find differences in a number of syntactic and semantic (and thematic, in their case) tasks. Unfortunately, it is not clear that these measures were reliable enough to accept the null results, and their sample sizes again are rather small (never exceeding 20 per group). Although the consistent failure to find higher-order tasks related to reading is intriguing, we are skeptical about concluding that higher-order processes are not influential in reading. The use of unstable measures in these experiments may have obscured the underlying pattern. Remedying this will be of considerable importance in our study.

Each of the three projects just described has stressed the role of letters and words in an information-processing framework of reading, while simultaneously minimizing the role of higher-order processes. In fact, Jackson and McClelland (in press) explicitly claim to

disconfirm at least the strongest versions of the psycholinguistic models -- "Surprising or not, our results do not support the views of some (e.g., Kolers, 1969; Smith, 1971) who have said that beyond the grade-school level, individual differences in reading ability are only differences in comprehension ability (in press)." While this statement may be an "over-interpretation" of Kolers' remarks, it nevertheless highlights the critical distinction between visual information processing and higher-order processing. As the next section on our research design will discuss in detail, we have included several higher-order tasks in this study. These were included both to contrast with visual information processes in their relationship to reading and to act as criteria for evaluating our visual information processing correlations with reading.

OVERVIEW OF THE RESEARCH

Because of the large number of tasks and analyses, it is important to set the stage before discussing our research. Consequently, this section briefly outlines our plan of attack for the rest of the paper.

The Experimental Procedures

The Method section contains a detailed description of each of the tasks that our subjects performed, together with the procedures involved in collecting the data. For purposes of clarity, we have separated discussion of the psychometric tests from discussion of the information-processing tasks. Thus, there are distinct method and procedure subsections for the two, with psychometric tests preceding information-processing tasks. A final subsection examines in some detail the characteristics of the subjects chosen for study.

It may help in reading the next section to keep in mind the rationale for our particular task selection. With regard to the psychometrics, we had available measures of verbal, spatial, and quantitative ability when selecting the subjects. To these, we added several other measures chosen to focus primarily on reading ability. As measures of reading comprehension, we used the Nelson-Denny Reading Test (1960) and a modified version of the Davis Reading Test (1962). As measures of reading speed, we used the Nelson-Denny test and a separate, time-

constrained passage - reading test devised in our laboratory by Joy Wright. Thus, we had two converging measures for both comprehension and speed of reading.

To permit separation of listening comprehension from reading comprehension, parallel forms of the modified Davis test were administered, one by tape recorder and one by booklet. In this way, we hoped to be able to partial out "general comprehension" from reading comprehension. By similar logic, our subjects were given the Raven Progressive Matrices Test (1965) as a measure of "general intelligence" which does not rely on verbal skills. The overall aim was to be able to single out reading-related processing from more general cognitive processing.

The choice of information-processing tasks was discussed earlier, with respect to precedents. All our tasks are quite basic to current views of reading and language -- visual search, span of apprehension, memory span, and name retrieval-matching. Each was performed with letters and with words, avoiding the confounding of unit and task. These were the tasks of primary interest, what we have been calling our measures of visual information processing.

There were also three higher-order tasks, again chosen from among the most familiar cognitive paradigms. First, there was a word-nonword lexical decision task. Although this could also potentially have served as a

measure of semantic decision making, it proved insensitive for this purpose. However, a second task provided a measure of semantic decision making. In this experiment, the subject had to verify the categorical relationship between two words. Finally, a syntactic task was included, the familiar picture-sentence verification task, in which the subject had to check the correspondence between a simple picture and a descriptive sentence. These are described in detail in the Method section, together with references to the appropriate literature. They were included to provide tasks which, while necessarily involving visual information processing, contained considerably more in the way of non-visual processing demands than the four primary tasks.

The Modes of Analysis

With a study of this size, a wide array of ways to present and analyze the data are possible, making clear presentation of the findings a formidable task. We have chosen to limit our analyses to several standard modes that we will use repeatedly. In terms of descriptive statistics, all of our reaction time data are represented by medians for correct trials only. We chose medians, as opposed to means, because reaction time distributions are typically skewed, and medians minimize this problem. The use of medians does have the drawback that they are generally less reliable than means, but this is a con-

servative error. Thus, medians are the inputs to all of the correlational analyses, and will be referred to simply as RTs henceforth. Error rate is also presented for each RT as it is introduced.

In terms of inferential statistics, we rely heavily on Pearson product-moment correlation coefficients, and on confidence intervals or standard errors about these. In every analysis, the reliability of the measures involved is demonstrated first. Typically, this involves split-half reliabilities over odd and even trials, although a few special cases will be pointed out in the Results section. Reliabilities are not corrected for test length, and may be seen as reasonably conservative estimates, as a result. To reduce the confusion that can occur in trying to interpret the sign of a correlation with so many different sorts of dependent variables, we have made all coefficients positive for positive relationships (e.g., lower RTs associated with higher ability, while algebraically a negative correlation has been assigned a positive value). For consistency, too, all of the confidence intervals used are associated with a $p < .05$.

Our major inferential tool is the Joreskog (1970) maximum likelihood method for the analysis of covariance structures. Frederiksen (Note 2) has also used this procedure in analyzing his data on adult readers, although

in a different way than we will use it. Where he uses it primarily to test his model, we will use it in a less theoretically complicated way simply to estimate correlational parameters and provide standard errors about those parameters. A brief discussion of the Joreskog method will help demonstrate its suitability for this sort of study.

The simplest application of the method, the one-trait model, is illustrated in Panel A of Figure 1. Assume that there is some hypothetical trait, $\underline{T}_1$, which is of interest. Several scores, $\underline{S}_1$ through $\underline{S}_n$, are available, each potentially providing a measure of the trait. The Joreskog method takes as input the correlation matrix of these scores. This matrix is submitted to a factor analysis, resulting in a set of factor loadings, $\underline{f}_1$ through $\underline{f}_n$, for those scores. Regenerating the correlation matrix on the basis of the factor analysis produces a new matrix, which can then be compared to the originally input matrix. A successful factor analysis will produce an exhaustive set of factors such that the two matrices are virtually identical, resulting in a nonsignificant χ^2. To the extent that the difference between the two matrices is large, χ^2 will also be large, demonstrating a less successful factor-analytic solution.

In short, the method determines how much of the variance among the scores is shared (i.e., reflects the common

trait) versus how much is error variance. Essentially, this is a sophisticated technique for determining the reliability of the variables. The only important assumption made is that the error variances themselves are not correlated; to the degree that they are, the individual parameter estimates will be inflated. This, of course, is a strong argument for using a set of converging measures, each of which is as distinct as possible from all of the other measures. The output of the analysis is the set of factor loadings (parameter estimates), $\underline{f}_1$ through $\underline{f}_n$, and the χ^2 value indicating how successfully these factor loadings account for the overall experimental variance.

By a straightforward extension of this logic, we arrive at the version of the Joreskog method most central to our analysis. This next level of complexity is shown in

- - - - - - - - - - - - - - -

Insert Figure 1 about here

- - - - - - - - - - - - - - -

Panel B of Figure 1. Here, the assumption is made that there are two traits, $\underline{T}_1$ and $\underline{T}_2$, each of which is tapped by only certain of the available scores. The important feature here is the possible relationship between the two traits themselves, defined by the parameter $\underline{r}_t$. It is at this parameter that hypothesis testing is aimed. For example, suppose we hypothesize that T_1 and T_2, com-

monly held to be separate traits, are in fact identical. This specific model can be tested by setting r_t equal to one and running the analysis. A nonsignificant χ^2 would be support for the model, although it would not necessarily rule out certain other models. However, a significant χ^2 would be evidence against this specific model.

The method can also be used in a less specific manner simply to estimate r_t with the error variance removed. This analysis additionally makes available a standard error estimate about r_t which is useful for hypothesis testing. In essence, the estimate of r_t includes a "correction for attenuation" that is superior to the usual correction because it relies on weaker assumptions. As anyone who has used it is aware, the standard correction for attenuation can produce a coefficient greater than one. The problem is that the standard method, while considering the variability in the correlation, does not take into account the variability in the reliability index. This assumption of "absolute" reliability is wrong. The Joreskog method considers variability in both measures, making it preferable.

The final level of application of the method that we will consider verges on Frederiksen's general model-testing use. The method is used almost as a general factor analysis, but with some correlations constrained, as a way to test particular hypotheses. Again, parameter

estimates and standard errors are obtained, this time together with tests of certain specific hypotheses. An illustration appears in Panel C of Figure 1. Suppose we are interested in the contribution of the letter unit to predicting reading speed versus reading comprehension. In addition, we wish to partial out choice reaction time from the RTs for letter tasks. The analysis would be set up as in Panel C and run twice, once with reading time and once with reading comprehension. In both cases, we would solve for r_t, and then compare them using the obtained confidence intervals. Subsequently, both reading measures could be inserted into an analysis, once unconstrained and once set equal to each other. If the unconstrained case with two free parameters does not offer a better fit than the one-parameter case, as defined by the difference in χ^2, this would be evidence that the two reading measures are overlapping.

To summarize, the analysis will rely on median correct RTs from the various tasks as the basic data. Correlations will be used to make inferences about the role of various processes, particularly visual processes, in reading. Tests of specific relationships will be conducted using the Joreskog method, which provides an estimate of r, a standard error, and a χ^2 value for the goodness of fit of a particular model. A χ^2 difference statistic can also be used to determine how much the fit is

reduced by constraining a given correlation to be equal to one (or zero), thereby losing a free parameter. This use of χ^2 differences is only possible in a setting such as ours where we are evaluating nested hypotheses.

The Discussion of Results

Because of the numerous analyses involved, we will report and discuss each set of findings in a subsection that is quite self-contained. The first section provides an overview of the nomothetic data, demonstrating successful replication of the various paradigms task by task. Where appropriate, we also point out the fits to the extant models. Reliabilities are reported for all variables in this section.

In the second section, the relationship between reading and the other psychometric measures is explored. Initially, the two reading measures, time and comprehension, are considered alone. This is followed by discussion of both reading measures as they relate first to listening comprehension and then to general intelligence. Each of these three analyses examines the specificity of the hypothetical reading ability trait.

The final two sections of the Results focus on the information processing tasks and their relationship to reading. The third section is aimed primarily at the visual information processing component, emphasizing the manipulation of the units over tasks. We examine letters

and words separately with respect to their role in both speed and comprehension of reading. The section concludes with a discussion of the "internal validity" of our letter and word measures. Finally, there is a shorter section on the higher-order tasks, their inter-relationship, and their relationship to reading. This section also explores the correlational pattern amongst all of our information processing tasks.

METHOD

Psychometric Measures

Reading: Comprehension and Time

(1) Nelson-Denny Reading Test. Form A of the comprehension and reading rate portion of the Nelson-Denny (1960) test was administered. Subjects were given 20 minutes to read eight selections and answer 4 multiple-choice questions at the end of each selection. While answering questions, subjects were allowed to refer back to a passage. If they had difficulty with a question, they were told to make a reasonable effort and then go on to the next question. Subjects used the Nelson-Denny self-marking answer sheet to record their responses.

The first minute of the test was used to determine reading rate. Subjects began reading the first passage and, at the end of one minute, were told to stop on the line they were reading and record that line number on their mark-sense forms. When this was completed, they returned to their reading.

(2) Reading Time. There were two speed-of-reading tests. The first test consisted of ten passages which were taken from a variety of magazines and text books. The passages were between 300 and 500 words in length, and corresponding lines numbers were printed beside each passage. A separate page containing three four-alternative multiple-choice comprehension questions followed

each passage. Subjects used mark-sense forms for recording reading rates and answers to the comprehension questions.

Subjects were instructed to read each passage in their normal way. At the end of 45 seconds, they were told to stop and record the line number that they had reached. Subjects were advised that it was in their best interest to mark line numbers accurately, because scores on the questions would be calculated only from material that had been read. If they claimed to have read more or less than they actually had, they would lower their scores. After marking the line number for a passage, subjects turned the page and answered the comprehension questions. They were given one minute before they were instructed to turn the page and read the next passage.

To converge on the results of this reading time test, we also administered the Minnesota Speed of Reading Test (Eurich, 1964). A standardized test, it consists of 36 short paragraphs, each containing an irrelevant phrase. The subject is scored on the number of such phrases correctly crossed out in 6 minutes.

All three reading-rate measures, the Nelson-Denny, our experimental measure, and the Minnesota, are usually expressed in terms of material read within a unit of time. This scale may be transformed into the inverse measure, the time to read a unit of material (i.e., msec.

per word). This latter measure has a simple theoretical interpretation in terms of the duration of mental processing. For this reason, all three reading rate measures will be transformed into reading-time measures, using as units milliseconds per word (cf. Table 2B).

(3) Reading and listening comprehension. The reading and listening comprehension tasks were based on identical materials. Forms A through D of the Davis Reading Test (1962) were used, with the original forms modified through deletion of passages and questions which contained references to tables or specific line numbers. Each modified form contained 12 passages and 45 questions. Forms A and D were made into separate booklets for the reading comprehension task. Forms B and C were tape-recorded for the listening task. The listening tapes were alternated with the reading booklets so that subjects were given a listening test, a reading test, a listening test, and a reading test. The order of tests was the same for all subjects.

For the reading task, subjects were instructed to read each passage carefully and then answer the questions without refering back to the material. Each passage in a booklet was followed by a page containing several five-alternative multiple-choice questions. Subjects recorded their responses on mark-sense forms, and they were told to guess only if they could rule out one or more of the five alternatives. Twenty minutes were alloted for each

of the two booklets.

During the listening task, subjects listened to a tape recording of a passage and then answered questions about it. Passages and questions were read in a normal speaking voice, at an average speed of 200 WPM. The tape recorder and amplifier were placed in the front of the room and subjects sat where they could listen comfortably. Each subject was given a test booklet containing only the five-alternative multiple-choice questions which corresponded to the passages. The questions were read to the subjects directly after each passage and were also written in the test booklet, otherwise the listening procedure was identical to the reading procedure. Each tape recording lasted 30 minutes and subjects listened to all of the passages and answered all of the questions at the same rate.

Intellectual Competence

(1) Raven Progressive Matrices. This is a standardized nonverbal intelligence test containing 36 test items using an identical format (Raven, 1965). Each item presents the subject with a pattern which can be thought of as a 3 x 3 matrix, with the entry corresponding to the third row-column missing. The task is to choose the missing element from among eight alternatives. The choice must be made by solving for the rule used in constructing elements across rows-columns. Items increase in diffi-

culty; subjects were told to be sure they had the correct answer before continuing to the next problem. They were given five minutes to complete ten practice problems, and 40 minutes to complete 36 test problems.

(2) Washington-Pre-College Test. Six scores from the Washington Pre-College Test (WPC) were used as psychometric measures in this study. The WPC (see Technical Manual, Note 6) is a standard college entrance test that the subject had taken prior to entering college. The six scores were:

(a) Vocabulary score is derived from a subtest on the WPC where students select the correct synonym for a given word from five choices.

(b) English-usage score is derived from a subtest which measures the student's ability in the use of grammar, punctuation, word choice, and capitalization.

(c) Spelling score is derived from a subtest which measures the student's ability to select a misspelled word from a choice of five words.

(d) Quantitative scores. Three scores from quantitative ability subtests are available: the Mathematics achievement, Quantitative skills and Applied mathematics subtests.

(e) Mechanical-reasoning score is derived from a subtest which measures the student's ability to answer questions about illustrations of mechanical systems.

(f) Spatial-ability score is derived from a subtest which

requires students to visualize how a two-dimensional figure would look in three dimensions if folded along certain lines.

General Procedure for Psychometrics. With the exception of the WPC-derived measures, both the Listening Comprehension Measures and the Reading Comprehension Measures, and all of the "Paper and Pencil" tests were group-administered in a large room. [The WPC was also group-administered, but was taken by the subjects as high-school juniors several years earlier.] The tests were timed, and subjects were instructed to work as quickly as possible without making errors. Each test required a booklet and a separate answer sheet.

Information-Processing Tasks

Overall procedure. As all the reaction-time tasks used the same general procedure, we describe this first. We then proceed to a detailed description of the specific tasks.

Subjects participated in a variety of tasks over seven-sessions. For the first four sessions of the study, subjects came for 50 minutes each day and performed the experimental reaction-time tasks. In the fifth session, subjects did a choice reaction time task, the span tasks, and a short reading-speed test in a 50-minute session. For these first five sessions, subjects participated in small groups of two to seven people. The sixth session

was 90 minutes long and consisted of the Nelson-Denny Reading Test and the Raven Matrices Test. The seventh session, several weeks later, lasted two hours, and subjects did the listening comprehension, reading comprehension and reading time tasks. Subjects participated in groups of one to ten for sessions six and seven of the study.

Because of the enormous number of subjects that would be required to counterbalance the order of 21 tasks, the schedule of experiments was fixed for all subjects. It was assumed that the order of experiments would not be a critical factor in this study. Table 1 shows the order in which tasks were presented.

- - - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - - -

Apparatus. A Nova 820 computer controlled stimulus presentation and response collection for the reaction-time tasks. Subjects were seated in individual sound-attenuating booths, each booth containing a response keyboard and an independently controlled Tektronix 604 cathode ray tube oscilloscope which presented the stimuli. The span tasks used the same apparatus as the reaction-time tasks with the exception that the Nova 820 computer was programmed only for stimulus presentation and was not used for response collection. Auditory stimuli for the

listening comprehension task were recorded and played back on a TEAC Model 1230 tape recorder using a Heathkit solid state amplifier, Model AA-18, and a Realistic Solo-4 speaker.

Reaction-Time Tasks. For all of the reaction-time tasks, subjects were instructed to work as quickly as they could without making errors. Accuracy was stressed more than speed, and subjects were told their data could not be used if they made more than ten errors per task (about 8% errors).

After each trial on a given task, subjects received "OK" or "NO" messages on their scopes (500 milliseconds in duration) informing them of the accuracy of their responses. After every eight trials, subjects saw their reaction times in milliseconds averaged over those eight trials, and the total number of errors they had made so far. This procedure was adopted to avoid having subjects "race" by attempting to respond faster than on the previous trial. Also, if they wished, subjects could relax for a moment in their booths after every eight trials. When they were ready to begin the next set of eight trials, they pressed any key on their response boards and the next trial began 250 milliseconds later. Thereafter, trials were spaced 500 milliseconds after trial feedback. Every trial was preceded by a warning dot which appeared for 250 milliseconds in the center of the

scope to inform subjects that stimuli were about to be presented. There was a 250-millisecond interval between the warning dot offset and the stimulus onset.

For all reaction-time tasks, there were two possible responses. Depending on the task, subjects used the right key on their response boards for a "SAME", "YES", or "RIGHT" response, and the left key for a "DIFFERENT", "NO", or "LEFT" response. Stimuli remained on the scopes until subjects responded. Stimulus conditions for each task were presented in a randomly permuted order; this order remained constant for all subjects, consistent with our decision to make procedures identical for all subjects.

With the exception of the choice-reaction-time task, subjects were given two sets of practice trials before each experiment. This gave the subjects an opportunity to see the procedure and ask questions if they did not understand it. The experimenter monitored errors and gave subjects any necessary help before the main part of the experiment began. The practice sessions reflected the same proportion of conditions as would be presented in the experimental tasks.

(1) Two-choice reaction time. For each trial of this task, subjects saw three stars to either the right or left of center on the scope. Subjects were instructed to press the RIGHT key with their right index fingers if the stars were on the right, and to press the LEFT key

with their left index fingers if the stars were on the left. There were eight practice trials and forty-eight regular trials. This task was performed first in each of the five reaction-time sessions because it refamiliarized subjects with the apparatus and gave a basic measure of simple key-pressing time for that session.

(2) Visual Search. This task required the subject to search for a target among a set of visually presented items (Sternberg & Scarborough, Note 7). On each trial of the letter-search and word-search tasks, a single target item was presented in the center of the scope for 250 milliseconds. It was then immediately replaced by a set of 2,3,4, or 5 items (the search set). If the target item appeared in the search set, subjects responded YES by pressing the right key. If it did not appear, they responded NO by pressing the left key. There were 32 practice trials and 128 test trials. The four sizes of search sets occurred equally often and were presented in a randomized order. On half of the trials of each set size, the target item occurred in the search set, with the target's serial position in the search set being counterbalanced.

For the letter version of this task, stimuli were constructed from the set of 24 lowercase letters in Appendix A (omitting i and q because of their visual similarity to other letters). On each trial, letter strings

of two to five items were randomly generated without replacement. For the word version of the task, the stimuli were constructed from a set of 24 common four-letter words (see Appendix A). All words were presented in lower-case, and word strings of two to five items were randomly generated without replacement on each trial.

(3) Matching. The matching task required subjects to determine whether two items had the same name, and was based on the task developed by Posner and Mitchell (1967). The stimulus set for the letter version of the matching task was made up of the 16 letters listed in Appendix A. These letters appeared in random pairs using all permutations of upper- and lower-case letters. Letter pairs were presented side by side in the center of the scope with one character space separating the letters. If the two letters were physically identical, or if they were physically different but had the same name, subjects were to respond SAME by pressing the right key. Otherwise, they responded DIFFERENT by pressing the left key. Of the 128 test pairs, 32 were physically the same, 32 had the same name, and 64 were different.

For the word version of the matching task, the stimulus items were 32 pairs of common, four-letter words (see Appendix A). The "different" test trials consisted of word pairs differing by one letter (e.g., SINK-WINK), with the changed letter balanced over position. The words

were presented in upper or lower case side by side with one character space separating them. As in the letter task, subjects pressed the right key if the two words had the same name, and the left key if they did not. Of 128 test pairs, 32 were physically identical (SINK SINK), 32 had the same name (SINK sink), and 64 had different names (SINK WINK; WINK sink). All word pairs occurred equally often in all appropriate conditions.

(4) Lexical decision. The lexical-decision task required subjects to judge whether an item was a word (Rubenstein, Lewis & Rubenstein, 1971). The stimulus set (cf. Appendix A) was made up of the lower-case word and nonword items used by Meyer, Schvaneveldt, and Ruddy (Note 8). Trials consisted of two items presented consecutively with each item requiring a response. Subjects were instructed to press the right key if an item was a word, and the left key if it was not a word. If subjects missed one or both items on a trial, the trial was counted wrong. There were 16 practice trials and 128 test trials. These 128 trials consisted of 16 pairs of associated words, 16 pairs of nonassociated words, 32 words followed by nonwords, 32 nonwords followed by words, and 32 pairs of nonwords. The sequencing of the trial types was random.

(5) Picture-sentence verification. This task required subjects to determine whether a sentence accurately described a picture (cf. Clark & Chase, 1972). The stimuli

consisted of two pictures, $\begin{smallmatrix}+*\end{smallmatrix}$ and $\begin{smallmatrix}*\\+\end{smallmatrix}$, and 16 lower-case descriptive sentences which varied in linguistic complexity (e.g., plus is above star; star is not below plus).

Subjects saw one of the two possible pictures for 1500 milliseconds, followed by one of the sentences. If the sentence correctly described the picture, the subjects were instructed to press the right key; otherwise, they were to press the left key. There were 16 practice trials and 128 test trials. The 128 trials consisted of 32 true affirmative (TA) trials where an affirmative sentence correctly described the picture ($\begin{smallmatrix}*\\+\end{smallmatrix}$ star is above plus), 32 false affirmative (FA) trials where an affirmative sentence incorrectly described the picture ($\begin{smallmatrix}+*\end{smallmatrix}$ plus is below star), 32 true negative (TN) trials where a negative sentence correctly described the picture ($\begin{smallmatrix}+*\end{smallmatrix}$ plus is not below star), and 32 false negative (FN) trials where a negative sentence incorrectly described the picture ($\begin{smallmatrix}*\\+\end{smallmatrix}$ star is not above plus).

It is possible to look at the four conditions in terms of whether or not the top item in a picture matches the first noun in a phrase. For example, a MATCH case would be $\begin{smallmatrix}+*\end{smallmatrix}$ plus is not above star. A mismatch case occurs when the top item in a picture does not match the first noun in a phrase, as in $\begin{smallmatrix}+*\end{smallmatrix}$ star is not below plus. Match and mismatch cases occurred equally often across

the four trial types.

(6) Semantic-decision. This task required subjects to determine whether a sentence was true, based on pre-existing semantic knowledge. The stimuli were patterned after the superset relations used by Collins and Quillian (1969), and appear in Appendix A. For the test trials, 40 three-level semantic hierarchies were used containing true and false sentences. Examples of the stimuli for this task are: A COBRA IS A COBRA (level zero), A COPPERHEAD IS A SNAKE (level one), A PYTHON IS A REPTILE (level two), and A VIPER IS A VOLCANO (false). The sentences were presented in random order one at a time in upper-case letters. Subjects were instructed to press the right key if a sentence was correct, and the left key if a sentence was not correct. There were 120 test trials and 16 practice trials.

(7) Inference making. This was the most complex of the reaction-time tasks, loosely based on the verification tasks in Anderson and Bower's (1973) work. Unfortunately, the pattern of main effects was not as anticipated; because we have no account for the results obtained, this task is not included in our analysis.

Span tasks general procedure. In the span tasks, the stimuli appeared either simultaneously (Span of Apprehension) or successively (Memory Span). Subjects were told to watch closely and, at the end of a trial, to

write the stimuli onto a response sheet in the order they were presented. Items recorded in the wrong order would be counted incorrect and no item would be repeated in a trial. A method of limits design was used with subjects told that the number of items per trial would increase by one after every three trials. The response sheet reflected this by having a blank space for every item to be presented on a given trial. Each task contained 15 trials. These were not timed tasks, and subjects could take as long as they wished to record the items. When subjects were ready to see the next trial, they pressed a key on their response boards and in 250 milliseconds a new trial began. Each trial was preceded by a fixation point which appeared for 250 milliseconds in the center of the scope followed 250 milliseconds later by the stimuli. The stimuli for both the letter and word versions of the span tasks were the same as those used in the search tasks.

Unlike the reaction-time tasks, subjects received no feedback during the test trials. Each span task began with a short session of three practice trials. At the end of the practice trials, subjects were shown the correct answers and asked if they had any questions.

(8) Memory span. Stimuli were presented one at a time in the center of the scope, with a 100-millisecond interval separating successive items. Each item was

displayed for 400 milliseconds. Four stimuli were presented for the three practice trials and for the first three test trials on both the letter and word versions of this task. With an item being added after every three test trials, subjects were seeing eight items per trial by the end of the experiment. They used separate response sheets for the letter and word versions of the task, but there was no difference between the sheets. Subjects were urged not to write on the sheets until a trial was completed.

(9) Span of apprehension. Stimuli were presented simultaneously, and remained on the scope for 100 milliseconds. Because of the difficulty of this task, two letters were presented for the three practice trials and first three test trials. Thus, subjects were seeing six letters by the end of the 15 test trials, one item added every three trials. One word was presented for the three practice trials and first three test trials of the word version. Subjects were seeing five words by the end of the 15 test trials.

SUBJECTS

As indicated previously, the sample was constructed by inviting participation from students who had taken the WPC prior to entering the university. The WPC provides a reading comprehension score based on the number of questions that can be answered about a passage after

a fixed time has been allocated to read it.

The range of reading comprehension scores of University of Washington students was divided into six groups. There were 15 subjects in each group with the exception of 1 group which had 16 subjects. This made a total of 91 subjects for the first portion of the study. Forty-six were males and 45 were females balanced as well as possible over the six groups. There was a follow-up session which occurred five months after the original six sessions, for which 75 subjects returned. Reading comprehension and listening comprehension scores were obtained in this seventh session. Thus, complete data were obtained for 67 subjects who completed all tasks. By examining their Nelson-Denny and Raven Progressive Matrices scores we can obtain an idea of the representativeness of the core 67-subject group. Figure 2 displays the frequency histogram of the Nelson-Denny reading test scores

- - - - - - - - - - - - - - -

Insert Figure 2 about here

- - - - - - - - - - - - - - -

plotted against the percentile norms from college graduates. Examination shows a reasonable fit to the expected rectangular distribution. A similar conclusion is reached with respect to intelligence when one inspects the Raven test scores in this manner. The sample appears to approximate the normal college student population.

RESULTS AND DISCUSSION

Nomothetic Results: Individual Experimental Tasks

We first review the outcomes of the individual information processing experiments. Our intention is two-fold -- to demonstrate a qualitative replication of the pattern of results that led to use of the paradigm and to locate quantitatively reliable measures for our individual differences application. Qualitative results are discussed in both nomothetic and individual-subject terms. Quantitative effects are primarily discussed in individual-subject terms with reliabilities calculated across subjects. A large collection of statistics concerning each measure is summarized in Tables 2A (reaction time tasks) and 2B (span tasks and psychometrics) including descriptive statistics, particularly standard deviations both within and between subjects, along with several reliability calculations. In general, all relevant results were replicated and were reliable (range from .70 to .96).

The data reported are median correct reaction times. Error rates and error reaction times are not analyzed because, as intended in our design, subjects made very few errors on any task.[2]

(1) Two-choice reaction time. This simple measure was collected first in each of the initial five sessions. Reaction time decreased steadily over sessions but also showed an increase in errors. The reliabilities also

improved over sessions (session-to-session: r = .67, .80, .75, .81), with an overall split-half reliability of .96.

(2) Visual search. Several alternative linear models have been used to describe search RT experiments, and our experiment replicates their general characteristics (cf. Sternberg,1966; Theios, Smith, Haviland, Traupmann & Moy, 1973). As Figure 3 shows, the data from this experiment fit a linear model well (letters: r^2 = .979; words: r^2 = .975, $d.f.$ = 2). A linear model was also

- - - - - - - - - - - - - - -

Insert Figure 3 about here

- - - - - - - - - - - - - - -

fit for individual subjects (letters: mean r^2 = .84; words: mean r^2 = .95, $d.f.$ = 2). In addition, large serial position effects were found, with the distance from the central fixation point accounting for as much overall variance as the set size [letters: r^2 = .84 (set size), .39 (serial position); words: r^2 = .49 (set size), .77 (serial position), $d.f.$ = 12]. It is worth pointing out that set size and serial position are not independent; in these experiments they are inherently correlated (r = .42, $d.f.$ = 12).

Gross performance measures were calculated for letters and words to permit comparison to other tasks. Their reliabilities were .94 and .86, respectively.

The difference between the gross RTs for letters and words (124 msec) also proved reliable (r = .73, n = 89 for all reliabilities). The linear models of set size were then calculated for individual subjects (n = 86). For letters, the slope and intercept estimates had reliabilities of .45 and .64, respectively; for words, these reliabilities were .72 and .49. The intercorrelations between the slope and intercept parameters were .12 for letters and .31 for words. These were calculated by correlating across two independent samples of data to remove any common error component.

(3) Matching. This experiment replicated the main result of a name identify (Aa) - physical identity (AA) difference (letters: 78 ± 12 msec; words: 100 ± 17 msec). The name identity-physical identity difference was calculated for each subject, but did not prove reliable (r <.30). As anticipated, the name-identity (Aa) and different-name (AB) conditions had approximately equal RTs (letters: 585,601; words: 834,839). The gross RTs were quite reliable (letters: r = .92; words: r = .94), as was the mean difference of 245 msec between the word and letter RTs (r = .86).

- - - - - - - - - - - - - - -

Insert Table 2A about here

- - - - - - - - - - - - - - -

(4) Lexical decision. Two main results were replicated.

First, the slower decision for nonwords as opposed to words was reliable for individual subjects (difference of 78 ± 10 msec, $\underline{r}$ = .65), as has been demonstrated frequently (e.g., Rubenstein et al., 1971). Although the semantic facilitation effect of associated words (NURSE-DOCTOR) was also replicated (e.g., Meyer et al., Note 8), it did not prove to be reliable over subjects (47 ± 12 msec, $\underline{r}$ <.3). However, the mean RTs were quite reliable -- calculated over blocks, the reliabilities were .83 for words, .90 for nonwords, and .85 overall.

(5) Picture-sentence verification. The original Clark and Chase (1972) experiment was replicated, as Figure 4 illustrates. A regression analysis shows that our data provide a reasonable fit to their model ($\underline{r}^2$ = .995, $\underline{d.f.}$ = 3). The reliabilities of the individual parameters

- - - - - - - - - - - - - - -

Insert Figure 4 about here

- - - - - - - - - - - - - - -

of the Clark and Chase model over 91 subjects were: Subject matching, .64; Markedness, .24; Negation, .74; and Falsification, .40. The mean RT was highly reliable ($\underline{r}$ = .94). [In addition, the relevant match cases were inspected for a fit to the Carpenter and Just (1975) model, which was quite good both nomothetically ($\underline{r}^2$ = .98, d.f. = 2) and on individual-subject data (mean $\underline{r}$ = .81).]

(6) Semantic decision. The basic pattern of hierarchical relationships (Collins & Quillian, 1969) was replicated.

The superordinate effect is clearly illustrated in Figure 5.

- - - - - - - - - - - - - - -

Insert Figure 5 about here

- - - - - - - - - - - - - - -

Unfortunately, while the critical difference between levels one and two is significant for the group data (225 ± 36 msec), the individual-subject difference scores were not reliable (r <.3). Still, the reliability of the mean RTs was .92 overall.

(7) Inference making. Because the overall pattern of group data was not readily interpretable, these data are not included in further analysis.

(8) Memory span. Little difference was observed between the number of letters reported (5.23 ± 0.17) and the number of words reported (5.13 ± 0.18). This was contrary to our expectation, based on the advantage usually found for letters (cf. Cavanaugh, 1972). Because the letter-word contrast for this task was rendered meaningless by

- - - - - - - - - - - - - - -

Insert Table 2B about here

- - - - - - - - - - - - - - -

this finding, the task is not included in the correlational analyses to follow. Intercorrelation between letter and word spans demonstrates a fair degree of reliability (r = .59).

(9) Span of apprehension. This experiment replicated

the long history of differences found in the tachistoscopic report of letters and words (Cattell, 1885). Although only 3.9 (±0.1) letters were reported in the letter version, 8.8 (±0.4) letters were reported in the word version (the equivalent of 2.2 words). Again, an acceptable lower bound on the task's reliability is demonstrated by the correlation between the letter and word versions (r = .50).

(10) Psychometric measures. The psychometric measures are described in Table 2B in similar detail to the information-processing measures also displayed. All psychometrics were reliable above .70, with reading time standing out at .92 (uncorrected for test length). All measures showed a wide range of variability, and were well away from both the ceiling and the floor.

Individual Differences Results: Correlational Analyses

Reading and the other Psychometric Measures

(1) Reading Time and Reading Comprehension. Both of our "experimental" reading measures proved to be highly correlated with relevant standardized tests. Our principal measure of reading comprehension, the modified Davis test, was highly correlated with both the Nelson-Denny (r = .72) and the WPC (comprehension r = .64; vocabulary, r = .68) standardized comprehension tests. Furthermore, our own Reading Time measure was highly correlated with two standardized reading time tests, the Nelson-Denny (r = .63)

and the Minnesota (r = .66)[3]. Each of these correlations confirms the validity of our experimental reading measures, the only measures we will discuss throughout subsequent analysis. For further consideration of the consistency of the reading measures, refer to the summary table of correlations in Appendix B.

If reading were a unitary skill, then one would expect our time and comprehension measures to be highly correlated. In fact, though, this correlation is only moderate (r = .53). The magnitude of this correlation is quite similar to that found in Blommers and Lindquist's (1944) time-constrained tests. Presumably, our correlation also reflects the time limitation on the comprehension test, and is thus somewhat inflated. Based on the literature (e.g., Blommers & Lindquist, 1944) the expected association between a timed reading-time test and an untimed comprehension test is r^2 = .10. The inflation is clear for our tests, where the comprehension test is time-limited, in that r^2 = .28. Even if our tests were perfectly correlated, the maximum r^2 would be .65, the rest being error.

Viewed in this way, the association between time and comprehension does seem to be moderate. However, because the association between time and comprehension is crucial in our upcoming analysis, a solid estimate of its magnitude is essential. With the Joreskog technique, the

reliability of the measures can be taken into account and used to provide a more representative view of the strength of association. For instance, consider the idea of two reading traits corresponding to time and comprehension. The Joreskog analysis estimates both the underlying relationship between the traits (r = .58 (.09), n = 72)[4], and the overall fit of the two-trait model (χ^2_1 = 2.64, p >.10). This nonsignificant chi-square supports the claim that reading time and reading comprehension cannot reflect a single trait. This is further emphasized by conducting a second Joreskog analysis, constraining the correlation between the two hypothetical reading traits to one. Not surprisingly, this unitary trait model fails (χ^2_2 = 39.09, p <.001). The difference between these two models is clearly significant (χ^2_1 = 36.55, p <.001). Reading ability is not a single trait.

(2) Listening Comprehension and the Reading Measures. The correlation between listening comprehension and reading comprehension was very high (r = .80, (.70, .87), n = 75). This is illustrated in the scatter plot in

- - - - - - - - - - - - - - -

Insert Figure 6 about here

- - - - - - - - - - - - - - -

Figure 6. Analysis of the distributions of scores from the two tests, and of the distribution of the residuals, shows no sign of skewness or heterogeneity of variance.

Although a test of linearity showed evidence for some nonlinearity, F (42,31) = 2.04, MS_e = 44, $p < .05$, improvement over a linear model was very slight and showed no systematic pattern (as Figure 6 demonstrates).

As foreshadowed in the introduction, the size of the reading-listening correlation is of considerable interest. If large, it limits the possible contribution to reading comprehension of any other processes, such as visual processes. The Joreskog analysis represents the relationship between reading comprehension and listening comprehension to be very strong indeed (r = .95 (.05), n = 75), χ^2_1 = 1.11, $p > .10$. Furthermore, the specific hypothesis that this correlation of listening and reading is equal to one cannot be rejected, χ^2_2 = 2.66, $p > .10$. The improvement in fit with the extra free parameter is not significant (χ^2_1 = 1.55, $p > .10$). Thus, the hypothesis that visual information processing contributes nothing unique to reading comprehension remains a viable one.

We are not, however, left without any place for visual processes in reading. This can be looked at from two perspectives. One might conclude that paragraph-comprehension tests are poor indices of the skills unique to reading. It is worth pointing out that the moderate correlation between reading time and listening comprehension (r = .52) is very close to that between reading time

and reading comprehension (r = .53). Perhaps time is a better measure of the skills unique to reading; if so, the moderate correlations with comprehension measures leave open the possibility of a large, uniquely visual contribution.

Alternatively, one might start from the premise that reading is not a unitary skill. While a comprehension test may measure primarily comprehension skill, other skills might be reflected in reading speed, and still other skills in other characteristics of reading (e.g., effort required to read a difficult versus an easy passage). Within this multivariate conception, the lack of a unique contribution of visual processes to comprehension is certainly reasonable, and is consistent with at least some theoretical positions (e.g., Smith, 1971).

The choice between these two views hinges on what one means by reading. Those holding a tight, univariate definition will find the present measures unacceptable, while those subscribing to a multivariate view are content that reading comprehension is well predicted by nonvisual tasks. We favor the latter because both accuracy and time are theoretically interesting aspects of reading.

(3) Reading and Intelligence. The relationship between reading and intelligence cannot be summarized by any single measure. Certainly, the correlations between

our two reading measures and several intelligence measures (contained in Table 3) are only part of the story. Table 3 is ordered roughly from tests that are most clearly verbal

- - - - - - - - - - - - - - -

Insert Table 3 about here

- - - - - - - - - - - - - - -

to those that are largely nonverbal; the respective correlations range from a high of .80 to a low of .08. Were a factor-analytic intelligence measure derived from these tests, it would show a moderate correlation with reading ability, as it would with spatial and quantitative ability. In principle, one could use this procedure to estimate the component of variation due uniquely to reading and not to general intelligence. However, this procedure makes no sense if one accepts, as we do (e.g., Hunt, 1978), a multivariate view of intelligence that includes components for verbal, spatial, and other abilities. Given the high correlations between verbal intelligence and reading comprehension at the top of Table 3, the two are simply not separable. Moreover, the virtual orthogonality of our reading measures with the Raven nonverbal intelligence measure demonstrates that this relationship is not worthwhile to pursue. Possibly, more specific concepts (and measures) of intelligence would reveal some interesting moderate correlations between "intelligence" and reading, but none are readily apparent. Given the observations

available, we believe that little benefit can be expected from pursuing the reading-intelligence relationship.

Reading and the Visual Information Processing Measures

(1) Reading and Letters. The correlations between the three letter tasks (span of apprehension, matching, and search) and the two principal reading measures, time and comprehension, were uniformly close to zero, as shown at the top of Table 4. A Joreskog analysis can be applied to determine an upper limit on this relationship, using the three-trait model described in the introduction. This procedure removes both error variability and systematic variability due to a two-choice RT ability.[5] We applied this model to the two measures involving letter RTs, search and match. The resulting estimates of the relation between letters and reading time (r = .03 (.45), n = 72); χ^2_6 = 9.36, p >.10) and between letters and reading comprehension (r = .01 (.11), n = 71); χ^2_6 = 7.20, p >.10) remain very low. These estimates are representative of several runs of this analysis, each producing different values. Clearly, the model of a unitary letter trait is not working, as the very large standard errors testify. The model simply does not converge well, and bolsters our contention that there is no unitary letter trait. When individual letters are used as stimuli, performance on the visual information processing tasks is unrelated to reading.

One might expect these low correlations involving letters to be improved if more selective measures were chosen to represent the search and matching tasks. This logic fails. The slopes and intercepts derived from the letter-search task were also not correlated with reading (time: r = -.01 and .11; comprehension: r = .06 and .10 for slopes and intercepts, respectively). Similarly, the three conditions of the matching task -- physical identity, name identity, and different name -- were also essentially uncorrelated with reading (time (n = 74): r = .12, .13, .23 respectively; comprehension (n = 75): r = -.02, .04, -.04, respectively). Obviously, more selective measures do not change our basic conclusion -- for letters, visual information processing is not related to reading.

Based on the literature, the search-measure correlations are not very surprising. Several authors (e.g., Katz & Wicklund, 1972) have failed to find differences between good and poor readers in visual search, but these were in the context of an extreme-groups design where the power of the measurement was uncertain. In a very careful study, Chiang and Atkinson (1976) did find a small relationship between visual search slope and verbal SAT (r = .35), but the small number of subjects (n = 30) and the failure of the correlation to reach statistical significance are grounds for skepticism. Furthermore, the

possibility that our near-zero correlations for slope are due to insufficient practice seems untenable. Hunt, McKee and Yantis (Note 9) also found near-zero correlations for their sample of 55 subjects, even after four days of practice (mean RT and slope by Nelson-Denny reading rate, r = .19,-.06; by Nelson-Denny comprehension, r = .20, .10). The consensus appears to be that visual search for letters has little predictive power with respect to reading.

For span of apprehension, the situation is not as straightforward. Jackson and McClelland (1975) pursued the classical result that good readers make fewer eye fixations per page than poor readers by correlating a letter span of apprehension measure with reading speed. They found a modest correlation (r = .48, n = 12, p = .06), which initially seems to conflict with our near-zero value. However, basing any conclusion on a nonsignificant correlation involving six subjects in each of two very extreme groups seems risky. We prefer to reserve judgment on this relationship.

Initially, there also seems to be some controversy surrounding the letter-matching task, especially the name-identity condition. Quite clearly, name identity RTs are correlated with general tests of verbal ability (see Hunt, 1978, for a review) -- for a college population like that in our study, these correlations are generally around .3. But Jackson and McClelland (in press) find

even larger correlations of name-identity RT with their two "effective reading speed" measures (r = .45, .49; n = 24). Once again, this conflicts with our near-zero correlation, and the caution concerning small samples and extreme groups designs must be reiterated. Hunt et al. (Note 9) offer a resolution of this verbal ability-reading ability discrepancy. In a group of well-practiced subjects (n = 55), they found the WPC-Verbal measure to be correlated with name-identity RT as usual, r = .40. However, correlations with both forms of the Nelson-Denny comprehension and time measures were low (r = .15 and .16, respectively). Perhaps letter matching is tapping some process unique to verbal ability tests? Regardless, the best summary appears to be that there is a limited, perhaps zero, correlation between letter tasks and reading, at least within the sample of three tasks reported here (span, search, and matching). That the unit is responsible, not the task, will become clear in our discussion of the same tasks using words as stimuli.

- - - - - - - - - - - - - - -

Insert Table 4 about here

- - - - - - - - - - - - - - -

(2) Reading and Words. The correlations between the three word tasks and the principal reading measures were in the small to moderate range, as shown in the bottom of Table 4. A Joreskog analysis identical to that done with letters provides an estimate of the word-reading

relationship. This relationship for reading time (r = .39 (.12), n = 74; χ^2_6 = 7.54, p >.10) is somewhat larger than for reading comprehension (r = .14 (.14), n = 73; χ^2_6 = 5.58, p >.10), although both are related. To evaluate the difference between the time and comprehension patterns, a second Joreskog model can be applied to simultaneously estimate both parameters. Unconstrained, this simultaneous model produces values for time (r = .38 (.11), n = 72) and for comprehension (r = .17 (.13), n = 72) that are very close to those found in the separate analyses. Constraining these two parameters to be equal produces a value of r = .29 (.11).

The reduction in fit from the two-parameter to the one-parameter model gives an idea of the significance of the distinct time and comprehension parameters. Constrained, χ^2_{16} = 16.80, p >.10, while unconstrained, χ^2_{15} = 13.12, p >.10 -- there is a marginally significant improvement with the addition of a single parameter (χ^2_1 = 3.68, .10 >p>.05). Unfortunately, neither of these tests is conclusive. Thus, while the available results point to the validity of the time-comprehension distinction, the critical evidence remains to be found.

The apparently stronger relationship between these word-processing tasks and reading time, as opposed to reading comprehension, fits with the multivariate view of reading introduced earlier. Imagine reading as a

serial process, and consider time and accuracy separately. Both early visual processing and later comprehension processing would contribute to the overall time measure. However, overall accuracy (the comprehension measure) would presumably be affected only by the comprehension process, the visual processes being essentially error-free. This two-stage notion is consistent with our results. There is a high correlation between listening and reading comprehension, moderate correlations between the time and comprehension measures, and evidence that visual information processing is more related to reading time than to reading comprehension. We will consider such a two-stage model again in the General Discussion.

A second interesting feature of these correlations is found in the contrast between letters and words; words appear to have a stronger relationship with reading than do letters. This can be demonstrated statistically in a way similar to that used in contrasting the reading measures themselves. A model that simultaneously estimates both the letter and word relationships to reading time shows a pattern very reminiscent of the separate analyses (letters: $\underline{r} = -.01$ (.13); words: $\underline{r} = .38$ (.11); $\chi^2_{13} = 17.12$, $\underline{p} > .10$). Constraining the word and letter parameters to be equal, the estimate becomes $\underline{r} = .29$ (.11), and the fit worsens ($\chi^2_{14} = 27.38$, $\underline{p} < .05$). The specific equality model must be rejected, as the difference between

chi-squares shows ($\chi^2_1 = 10.26$, $p < .01$, $n = 73$). Words predict reading time better than do letters.

The literature contains a number of isolated pieces of evidence that word visual information processing tasks are related to reading. Although the fact that these are generally small-sample, extreme-groups studies makes estimation of the strength of the relationship difficult, the recurring pattern is reassuring. For example, variations on the span of apprehension task with words have been reported to be related to reading in sentence contexts (Jackson & McClelland, 1975). Also, Perfetti et al. (Note 3) have shown that visual search with words predicts reading. The current study goes beyond the literature in three ways. First, we are able to provide an estimate of the strength of the relationship between words and reading. Second, we provide evidence that the relation of words to reading time is somewhat greater than to reading comprehension. Third, the use of letters and words in the same tasks permits a direct comparison of the two units.

The contrast between letters and words has several implications. Apparently, materials make more difference than do tasks. If the same operations were performed on both these kinds of materials, this would not be expected. Word visual information processing cannot be entirely a concatenation of letter-handling processes, because the units are not interchangeable within a task. Further

evidence regarding this point can be obtained from the direct correlation between letters and words, which we will take up following a consideration of the internal correlations within these two unit types over tasks.

(3) Internal Consistency of Letter and Word Tasks. Any effort to determine the relationship between visual information processing and reading depends on the validity of the chosen measures of visual information processing. The degree of convergence achieved will be revealed by the intercorrelations within the letter and word measures. If the letter measures from different tasks are not correlated with each other, then by inference they do not share the same processes. Furthermore, if this level of consistency fails, the likelihood of meaningful correlations with reading ability is small.

For letters, the correlations between span of apprehension and each of the two RT tasks is low (search: r = .21; matching: r = .21). Although the correlation between search and matching is large (r = .73), in part it must surely reflect the shared response requirements of two-choice RT. A Joreskog analysis can separate out this contribution, using even and odd trials as separate inputs for each task. Assume that there are three traits corresponding to the matching, search, and two-choice RT tasks, and that the correlations between the choice trait and the other two traits are fixed to zero. This

model estimates a moderate relationship between search and matching (r = .63 (.07), n = 89; χ^2_4 = 12.62, p <.05). Thus, the search and matching tasks converge on something beside two-choice RT, although there is little evidence that they are linked to span of apprehension. In conclusion, the validity of the letter unit as a uniform trait is in doubt, as suggested earlier.

The same validity analysis applied to words leads to more positive results. The span measure is moderately related to each of the RT tasks (search: r = .37; matching: r = .42), which are again highly correlated themselves (r = .77). A Joreskog analysis confirms the strong relationship between word search and matching, with two-choice RT as a separate trait (r = .74 (.06), n = 89; χ^2_4 = 1.85, p >.10). These results are somewhat more encouraging, and reinforce our earlier suggestion of a word trait.

(4) The Relation between Letter and Word Tasks. Table 5 presents the intercorrelations between the letter and word tasks, and indicates something about the extent to

- - - - - - - - - - - - - - -

Insert Table 5 about here

- - - - - - - - - - - - - - -

which processes are shared over letters and words. Not surprisingly, the largest correlations are those on the diagonal -- the same task with different stimuli. Cor-

relations off the diagonal share neither task nor stimulus material. These suggest that, when shared task-specific variation is not a factor, the relationship between letters and words is quite low. To estimate this value, the optimal Joreskog analysis would seem to be one in which all task variables are removed, leaving only the letter-word relationship. Unfortunately, such a model is not defined for our case because the ratio of free parameters to data points is too low. To answer the question, then, we conducted a Joreskog analysis with just the two RT tasks, search and match, separated for letters versus words, and with two-choice RT as a trait. Although this model appears to offer a fair fit at first look ($\underline{r}$ = .72 (.08), $\underline{n}$ = 86; χ^2_4 = 5.88, $\underline{p}$ >10), it is not that simple. If one correlates span and the search-match composite across units, both correlations are very low relative to the above RT-to-RT estimate (letter span to word RT composite, $\underline{r}$ = .26; word span to letter RT composite, $\underline{r}$ = .19). These should be of the same magnitude as the Joreskog estimate, but they are not. Perhaps there is some additional relationship between the RT tasks that is inflating the Joreskog estimate, but we cannot be certain without the appropriate covariates. Also, the lack of coherence in the letter tasks may be influencing this pattern. We conclude that while there is some common variance between the two units, it is

probably not very large. It would appear that processes unique to letters or to words dominate the processing of their respective units.

(5) Visual Information Processing and Reading Reconsidered. One final and most conservative test can be made of the relationship between word tasks and reading. The idea is to set up a Joreskog analysis with span of apprehension, search, and matching as a composite variable representing the hypothetical word trait opposite each of the reading measures in turn. In such an analysis, any variability not shared by these word measures will be treated as error. This results in an increased range of possible values for the relation between words and reading. Essentially, the procedure serves as a "correction for invalidity" in the same sense that the usual correction for attenuation serves to correct for the error variability between parallel tests. The goal in using this estimate is to provide an upper bound on the size of the word-reading relationship. The model produces a moderately large estimate with reading time (r = .52 (.13), n = 74; χ^2_1 = 1.08, p >.10), and a somewhat smaller estimate with reading comprehension (r = .29 (.16), n = 73; χ^2_1 = 1.05, p >.10). Still, both of these estimates are considerably larger than one might have anticipated on the basis of simple correlations, which ignore within-subject error. These estimates, with their standard errors, should provide some reassurance

that the simple correlational relationship between a word trait and reading is, if anything, an underestimate of its strength.

It may help to reiterate our results thus far before discussing the results from the higher-order tasks. When letters versus words are used as the stimuli in the same visual information processing tasks, a different pattern of correlations with reading ability emerges. Letter tasks show little correlation to either reading speed or comprehension; indeed, we have demonstrated that the hypothesis of no relationship remains a possibility. In contrast, word tasks show a small but consistent relationship with reading. By using Joreskog's structural analysis of covariance matrices, we have demonstrated that a word-reading relationship independent of task-specific variance exists. With corrections for reliability and validity, estimates for this relationship with reading time, and to a lesser extent reading comprehension, are quite high.

Higher-Order Information Processing Tasks

(1) Relationship to Reading. The three higher-order tasks all showed small to moderate correlations with reading ability, as displayed in Table 6 separately for time and comprehension. While it is tempting to think of

- - - - - - - - - - - - - - -

Insert Table 6 about here

- - - - - - - - - - - - - - -

these tasks as representing a higher level stimulus unit such as sentences, this almost certainly oversteps the capabilities of our design. We will simply treat them as more complex, without attempting to specify the dimensions of complexity. A Joreskog analysis of these as a set of converging measures (with two-choice RT removed) estimates their relationships to reading time (r = .47 (.10), n = 71; χ^2_{10} = 11.55, p >.10) and to reading comprehension (r = .41 (.11), n = 70; χ^2_{10} = 15.88, p >.10) as moderate. These gross RT measures can be supplemented by derived measures for two of the three higher-order tasks. In the lexical decision task, the nonword-word difference shows a similar small to moderate correlation with reading time (r = .44) and with reading comprehension (r = .31, n = 72 for both). The other derived measure, the negative-affirmative difference in picture-sentence verification, shows a weaker relationship with both reading time (r = .16) and reading comprehension (r = .28, n = 75 for both). In summary, virtually all of the reliable higher-order measures showed some relationship with our two measures of reading ability.

The principal reason for presenting these particular results is to contrast the higher-order with the simpler information processing tasks. The higher-order tasks are at least as highly, and perhaps more highly, correlated with reading as the simple visual tasks that use words

as stimuli. Previous studies have revealed similar results for lexical decision (Frederiksen, Note 1) and for sentence-picture verification (MacLeod et al., 1978), so that we appear to be on firm ground. However, Perfetti and Lesgold (1977) have argued precisely the converse -- they argue that simple processes play a larger role than do higher-order processes in reading. Indeed, their data do not reveal reliable patterns with several discourse comprehension tasks as a function of reading, but these failures may well have been due to lack of statistical power. We would be particularly concerned about reliability in the individual scores (despite orderly group data) because of the apparent greater difficulty in obtaining reliable higher-order tasks. Furthermore, it seems more reasonable to expect that tasks more closely related to normal reading would be more highly correlated with the overall process, as appears to be the case in our study.

(2) Relations Among Higher-Order Tasks. Unlike the visual information processing tasks used with letters and words, which were selected for their consistency, the higher-order tasks were selected for their diversity. They were intended to represent a range of possible lexical (lexical decision), syntactic (picture-sentence verification), and semantic (semantic decision) processes that might follow initial visual processing. Because we

have no specific predictions about their interrelations, this brief section is simply exploratory. The first thing to point out is the simple correlations. The tasks are all highly related: lexical decision to picture-sentence (r = .55), picture-sentence to semantic decision (r = .64), and semantic decision to lexical decision (r = .84). This pattern is confirmed by a set of three Joreskog analyses removing error and the two-choice RT contribution from each pairwise combination of the three tasks. This was done simply to obtain estimates, and not to test specific hypotheses as previous analyses have. The results reinforce the simple correlational analysis: lexical decision with picture-sentence (r = .45 (.10), n = 88; χ^2_4 = 0.69, p >.10), picture-sentence with semantic decision (r = .54 (.08), n = 91; χ^2_4 = 1.47, p >.10), and semantic decision with lexical decision (r = .81 (.05), n = 88; χ^2_4 = 1.23, p >.10). As can be seen from either the simple correlations or the Joreskog estimates, lexical decision and semantic decision stand out as being very closely related. Considerably less related was the picture-sentence verification task to either of these. It might be argued that what we are seeing here is some effect of requiring versus not requiring access to meaning. However, this does not seem reasonable, given the lack of a systematic correlational pattern for semantic decision or lexical decision with verbal psychometric measures. Resolution of this

pattern must await a more specific theoretical framework.

(3) Relation of Visual Information Processing Tasks to Higher-Order Tasks. As letter and word tasks have been contrasted in their relationship with reading, they can now be contrasted in their relationship with the higher-order tasks. The pattern of these individual correlations, contained in Appendix B, is fairly heterogeneous. They may be summarized most simply by the correlations between composite variables representing the letter, word, and higher-order tasks. These composites were formed by summing the standardized scores for each of the component tasks. Simply inspecting Table 7 makes it clear that words

- - - - - - - - - - - - - -

Insert Table 7 about here

- - - - - - - - - - - - - -

show a greater correlation to higher-order tasks than do letters. More analytically, Joreskog analysis can be applied here, keeping in mind that the higher-order tasks were never selected to tap a single trait. In any case, the relation estimated for letters (r = .43 (.12), n = 86; χ^2_8 = 20.83, p <.01) is clearly smaller than that for words (r = .81 (.06), n = 88; χ^2_8 = 8.40, p >.10). This analysis reinforces the earlier conclusions about the letter tasks. Letter tasks are not related to reading speed or comprehension, show limited internal consistency, and seem only slightly related to other information processing tasks.

GENERAL DISCUSSION

In this paper, our goal has been to provide information concerning the question "Is visual information processing related to reading?" The empirical correlations between various converging measures have been discussed in some detail, so we may now consider the implications of those results. In particular, implications of the listening-reading and letter-word contrasts will be our focus. Then, more general questions will be raised concerning theories of reading -- in particular, information processing theories of reading. This will lead to a concluding section developing one possible information processing model.

Consider first the listening-reading contrast. Arguments about their similarity seem to bounce between two extremes. On the one hand, it has been asserted that a brief, modality-specific encoding stage is followed by uniform linguistic processing regardless of the modality. On the other hand, it has been asserted that the differences between listening and reading lie at a deeper level than simply ear versus eye. Typically, three arguments are raised for their dissimilarity (cf. Mattingly, 1972):

(1) The nature of the code that must be broken is different. Speech is highly redundant, while visual representations are more compactly encoded;

(2) The usual listening situation is full of additional

sources of information and opportunities for interaction, all of which are missing in reading;

(3) The material conveyed by speech is not usually as complex as written prose. The differences between spoken lectures and written texts are frequently invoked as examples.

These differences, especially the latter two, imply that the nature of the message is different for the reader as compared to the listener. And if the message is different, certainly the processing of it must be different as well.

In our comparison of reading and listening, considerable effort was expended to make the messages identical. By using interchangeable materials and taped auditory presentation, the differences in content were minimized. Virtually all that remained was the difference in modality. It is under these conditions that we find the very close similarity between reading and listening comprehension abilities. The implication is that modality-specific processing, given the same message, does not contribute to comprehension ability. Of course, this allows for the possibility that the gist of the above extreme arguments is correct. Ordinarily, the differences in content between listening and reading may well be large; however, when controlled, the differences in processing may be minimal.

Consider next the letter-word contrast. Throughout

this paper, we have reiterated the observation that the ability to process individual letters is not related to reading ability. This is certainly not predicted by those who emphasize the role of letters in word recognition (e.g., Gough, 1972), but it is predicted by those who emphasize the unique contribution of "wordness" to word recognition (e.g., Reicher, 1969). Thus, our study adds another piece of evidence to the existing case against letter-oriented models. However, between these extreme views of word recognition, many intermediate possibilities exist (cf. Smith & Spoehr, 1974). Perhaps individual differences research can aid in distinguishing these as well. For example, Mason and Katz (1976) and Frederiksen (Note 1) have demonstrated that there is some relationship between the processing of redundant letter strings and reading ability. If performance on these related letter strings predicted performance on words, the notion of common underlying processes would be supported. Just as interesting would be a failure to show a close relationship, which would necessitate a continued search for an explanation of "wordness".

Can there be an information-processing theory of reading? The heart of information processing is the discovery of distinct processes which operate on information. Unfortunately, efforts to specify such processes in reading have not been very successful. Indeed,

the alternative of analysis-by-synthesis (or other highly interactive model of processing) is at least equally popular (for a review, see Gibson & Levin, 1975). The present research has been directed at uncovering the differences among tasks that might identify the distinct processes proposed on the basis of an information-processing analysis. Two of our principal results are the observation of a difference between speed and comprehension in reading and between letters and words as the unit in visual information processing. These observed differences suggest a possible rudimentary division of the component processes. As a first step, comprehension and time must subsume different processes. While these processes may be far from distinct, it is easiest to start with the simplest hypotheses and reject those.

Our modeling of reading ability begins with the fundamental observation that reading is a multivariate behavior. Speed and comprehension do not reflect the same processes. Consequently, we will entertain a two-component model. Furthermore, for simplicity, we will assume that these components are independent. One of these components must be a general comprehension facility, identified on the basis of the common processes behind listening and reading comprehension. Their high correlation strongly supports the need for such a stage. But listening is nonvisual, so the obvious candidate for the other component

in our model is visual processing. In tests of reading ability, only the comprehension component would relate to an individual's comprehension score, but both components would contribute to overall reading time. This is logically demanded by the intermediate correlations between the time and comprehension measures, coupled with the assumption that the components are independent.

The visual processing component remains to be identified. To fit the model, the representatives for this construct should be correlated with reading time and not with reading comprehension. The letter measures clearly do not fit these expectations. There is more hope for the word measures, though; they are moderately correlated with reading time and less correlated with comprehension. The higher-order tasks also show correlations with both of the reading measures, which can be taken in two ways. First, this may be simply evidence against the independence-of-components assumption. Alternatively, the use of timed comprehension tests may be inflating the correlation between reading comprehension and the word tasks just as it inflates the reading time-reading comprehension correlation. We suggest that, until the latter explanation is empirically invalidated, the independence assumption should not be abandoned.

The identification of the precise visual processes of this model must be left incomplete. While one candi-

date -- letters -- can be rejected and another -- words -- looks promising, one study cannot resolve the issue. Still, the choice of word processes is consistent with the existing literature relevant to this model. For instance, in a large psychometric study, Holmes (1954) tested a set of measures as predictors of reading speed. After discarding measures of comprehension and verbal intelligence, the next best predictor was a word discrimination measure. What makes this more interesting is an examination of the measures that failed to compete as predictors. These included such measures as grammatical and perceptual abilities, along with eye-movement parameters. Although we cannot accept the crucial role of visual-word processing on the strength of a single study, the pattern is suggestive.

This paper has addressed the question of whether visual information processing is related to reading. A simple, two-stage model has been developed to capture the component abilities underlying reading. The superordinate feature of this model is reliance on a multivariate view of reading ability. The model provides an initial global account of the different processes involved in reading comprehension versus reading speed. Within this account, the model offers some testable hypotheses about the nature of the specific visual and general comprehension processes in reading, and their

relationship to speed versus comprehension. It is hoped that this framework will be useful in understanding the nature of reading in general and especially in exploring variations in reading ability.

REFERENCE NOTES

Note 1. Frederiksen, J. R. Assessment of perceptual, decoding, and lexical skills and their relation to reading proficiency. Bolt, Beranek and Newman, Inc., Report No. 3756, Technical Report No. 1, Cambridge, MA, January 1978.

Note 2. Frederiksen, J. R. A chronometric study of component skills in reading. Bolt, Beranek and Newman, Inc., Report No. 3757, Technical Report No. 2, Cambridge, MA, January 1978.

Note 3. Perfetti, C. A., Bell, L. C., Hogaboam, T. W., & Goldman, S. R. Visual processing speed and reading skill. Paper presented at the meeting of the Psychonomic Society, Washington, D. C., November 1977.

Note 4. Adams, M. J. Models of word recognition. Center for the Study of Reading, Technical Report No. 107, Urbana, Illinois, October 1978.

Note 5. Smiley, S. S., Oakley, D. D., Worthen, D., Campione, J. C., & Brown, A. L. Recall of thematically relevant material by adolescent good and poor readers as a function of written versus oral presentation. Center for the Study of Reading, Technical Report No. 23, Urbana, Illinois, March 1977.

Note 6. University of Washington Assessment Center. Technical manual: Washington pre-college test battery. Seattle, 1977.

Note 7. Sternberg, S., & Scarborough, D. L. Parallel testing of stimuli in visual search. Paper presented at the international symposium on Visual Information Processing and Control of Motor Activity, Sofia, Bulgaria, July 1969.

Note 8. Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. Activation of lexical memory. Paper presented at the meeting of the Psychonomic Society, St. Louis, Missouri, November 1972.

Note 9. Hunt, E., McKee, C., & Yantis, S. Comparison of automated processing paradigms. Paper presented at the meeting of the Psychonomic Society, San Antonio, TX, November 1978.

REFERENCES

Anderson, J. R., & Bower, G. H. Human associative memory. Washington, D. C.: V. H. Winston, 1973.

Blommers, P., & Lindquist, E. F. Rate of comprehension of reading: Its measurement and its relation to comprehension. Journal of Educational Psychology, 1944, 35, 449-473.

Carpenter, P. A., & Just, M. A. Sentence comprehension: A psycholinguistic processing model of verification. Psychological Review, 1975, 82, 45-73.

Cattell, J. M. Ueber die Zeit der Erkennung und Benennung von Schriftzeichen, Bildern, und Farben. Philosophische Studien, 1885, 2, 635-650. Reprinted in translation in A. T. Poffenberger (Ed.), James McKeen Cattell: Man of Science. York, PA: Science Press, 1947.

Cavanaugh, J. P. Relation between the immediate memory span and the mrmory search rate. Psychological Review, 1972, 79, 525-530.

Chiang, A., & Atkinson, R. C. Individual differences and interrelationships among a set of cognitive skills. Memory and Cognition, 1976, 4, 661-672.

Clark, H. H., & Chase, W. G. On the process of comparing sentences against pictures. Cognitive Psychology, 1972, 3, 472-517.

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. Journal of Verbal Learning and

Verbal Behavior, 1969, 8, 240-247.

Davis, F. B., & Davis, C. C. Davis reading test. New York: The Psychological Corporation, 1962..

Estes, W. K. The locus of inferential and perceptual processes in letter recognition. Journal of Experimental Psychology: General, 1975, 104, 122-145.

Eurich, A. C. Minnesota speed of reading test. Minneapolis: University of Minnesota Press, 1964.

Gibson, E. J., & Levin, H. The psychology of reading. Cambridge, MA: The M. I. T. Press, 1975.

Gough, P. B. One second of reading. In J. F. Kavanagh and I. G. Mattingly (Eds.), Language by ear and by eye. Cambridge, MA: The M. I. T. Press, 1972.

Guthrie, J. T., & Tyler, S. J. Psycholinguistic processing in reading and listening among good and poor readers. Journal of Reading Behavior, 1976, 8, 415-426.

Holmes, J. A. Factors underlying major reading disabilities at the college level. Genetic Psychology Monographs, 1954, 49, 3-95.

Hunt, E. Mechanics of verbal ability. Psychological Review, 1978, 85, 109-130.

Jackson, M. D., & McClelland, J. L. Sensory and cognitive determinants of reading speed. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 565-574.

Jackson, M. D., & McClelland, J. L. Processing determinants of reading speed. Journal of Experimental Psychology: General, 1979, in press.

Jöreskog, K. G. A general method for analysis of covariance structures. Biometrika, 1970, 57, 239-251.

Katz, L., & Wicklund, D. A. Letter scanning rate for good and poor readers in grades two and six. Journal of Educational Psychology, 1972, 63, 363-367.

Kavanagh, J. F., & Mattingly, I. G. Language by ear and by eye: The relationships between speech and reading. Cambridge, MA: The M. I. T. Press, 1972.

Kolers, P. A. Reading is only incidentally visual. In K. S. Goodman and J. Fleming (Eds.), Psycholinguistics and the teaching of reading. Newark, DE: International Reading Association, 1969.

Mackworth, J. F. Some models of the reading process: Learners and skilled readers. Reading Research Quarterly, 1972, 7, 701-733.

MacLeod, C. M., Hunt, E. B., & Mathews, N. N. Individual differences in the verification of sentence-picture relationships. Journal of Verbal Learning and Verbal Behavior, 1978, 17, 493-507.

Mason, M., & Katz, L. Visual processing of nonlinguistic strings: Redundancy effects and reading ability. Journal of Experimental Psychology: General, 1976, 105, 338-348.

Mattingly, I. G. Reading, the linguistic process, and linguistic awareness. In J. F. Kavanagh and I. G. Mattingly (Eds.), Language by ear and by eye. Cambridge, MA: The M. I. T. Press, 1972.

McConkie, G. W., Rayner, K., & Wilson, S. J. Experimental manipulation of reading strategies. Journal of Educational Psychology, 1973, 65, 1-8.

Nelson, M. J., & Denny, E. C. The Nelson-Denny reading test. Boston: Houghton Mifflin Co., 1960.

Perfetti, C. A., & Hogaboam, T. The relationship between single word decoding and reading comprehension skill. Journal of Educational Psychology, 1975, 67, 461-469.

Perfetti, C. A., & Lesgold, A. M. Discourse comprehension and sources of individual differences. In M. A. Just and P. A. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, NJ: Erlbaum, 1977.

Posner, M. I., & Mitchell, R. F. Chronometric analysis of classification. Psychological Review, 1967, 74, 392-409.

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. Retention of visual and name codes of single letters. Journal of Experimental Psychology Monograph, 1969, 79 (1, Pt. 2).

Raven, J. C. Advanced progressive matrices, Sets I and II. London: Lewis, 1965.

Rayner, K. The perceptual span and peripheral cues in reading. Cognitive Psychology, 1975, 7, 65-81.

Reicher, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. Journal of Experimental Psychology, 1969, 81, 275-280.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. Evidence for phonemic recoding in visual word recognition. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 645-657.

Smith, E. E., & Spoehr, K. T. The perception of printed English: A theoretical perspective. In B. H. Kantowitz (Ed.), Human information processing: Tutorials in performance and cognition. Potomac, MD: Erlbaum, 1974.

Smith, F. Understanding reading. New York: Holt, Rinehart, & Winston, 1971.

Sternberg, S. High-speed scanning in human memory. Science, 1966, 153, 652-654.

Sticht, T. G. Learning by listening. In J. B. Carroll and R. O. Freedle (Ed.), Language comprehension and the acquisition of knowledge. Washington, D. C.: V. H. Winston, 1972.

Theios, J., Smith, P. G., Haviland, S. E., Traupmann, J., & Moy, M. C. Memory scanning as a serial self-terminating process. Journal of Experimental Psychology, 1973, 97, 323-336.

Tinker, M. A. Speed versus comprehension in reading as affected by level of difficulty. Journal of Edu-

cational Psychology, 1939, 30, 81-94.

Underwood, B. J. Individual differences as a crucible in theory construction. American Psychologist, 1975, 30, 128-134.

FOOTNOTES

This research was supported by National Institute of Education grant G77-0012 to the University of Washington. We are grateful to the following people for their contributions: Steve Poltrock for early discussions of the ideas, Marcy Lansman for critical comments on an earlier draft, Joy Wright for compiling the reading speed test, and Colene McKee for assistance with programming and data analysis. We also thank David E. Meyer for providing his materials for the lexical decision task, and Frank Andrews for statistical advice.

Requests for reprints may be sent to Earl Hunt, Department of Psychology, NI-25, University of Washington, Seattle, Washington 98195.

[1]An example of this latter approach illustrates its application. One might ask whether studies of tachistoscopic recognition are at all relevant to the study of reading. Since the early studies of Cattell (1885), a large body of literature has been built up concerning factors involved in tachistoscopic word recognition. Only recently has another experimental method, Rayner's (1975) "moving window" technique, converged on these results. Although his paradigm is not equivalent to normal reading, his results do suggest an analogy between tachistoscopic studies and more naturalistic reading studies.

[2]Two comments are necessary concerning speed-accuracy tradeoffs in the reaction-time experiments. First, nomothetic tradeoffs are indicated by a negative relationship between RT and accuracy over conditions. In our study, the main measures of interest are all single RT measures, so the usual tradeoff concern is not relevant. However, in cases where we do discuss differences in RT, accuracy follows the same pattern (cf. Table 2).

A second aspect of the speed-accuracy tradeoff is unique to individual differences research. This is the possibility that a correlation is due to the accuracy criterion and not the "true" mental processing duration. For example, if letter-matching time is correlated to reading, it may be due either to good readers being able to match letters faster or merely guessing a letter match or mismatch earlier. In the latter case, one would expect fast readers to show more errors. If the error-rate correlation is the negative of the RT correlation (each with the same variable), there may be an accuracy bias. Jackson and McClelland (in press) attempted to tackle this problem with partial correlations, the goal being to remove the variance in RT attributable to individual differences in accuracy. However, in our study (and possibly in theirs), the error rates are low and unreliable. For instance, only the picture-sentence

verification task had reliable errors in our study, and these were positively correlated with both correct RT and the other relevant measures. While it is true that the correlations between accuracy and RT tended to be negative in most of the other tasks (suggesting within-task speed-accuracy tradeoffs), this is independent of the critical correlation of an accuracy bias with the psychometrics.

An alternative approach to describing accuracy bias is to look for a general effect over tasks, assuming that subjects are consistent in applying their biases. Thus, a composite error measure was obtained from each of the RT measures by summing standardized scores. This composite error rate showed consistent negative correlations with RT in the -.25 to -.45 range. The consistency of this composite error rate's correlations demonstrates the existence of a general accuracy-bias trait. How does this trait correlate with the psychometrics? These correlations were essentially zero for all of the reading measures. Thus, the effect of the accuracy bias is only to inflate the intercorrelations between RT tasks, and not to affect the correlations of RTs with psychometrics. This finding eliminates the possibility of a general accuracy bias affecting psychometric correlations. However, there remains the possibility of an effect specific to certain individual tasks. This cannot be assessed in our study, or perhaps in any other study where

subjects are strongly discouraged from making errors and, hence, error rates are very low (under 5%). A more reliable measure of error rate would be necessary.·

[3]The distribution of reading time scores was fairly skewed (skew = .9). Each of two different transformations was applied in an attempt to achieve approximate normality. The correlation with reading comprehension increased only slightly for both the logarithmic (r = .55) and the speed (r = .57) transformations. Basically, while both resulted in symmetrical distributions, they had little overall effect on the correlational pattern. However, because of this skewness, we examined the character of the time-comprehension correlation in more detail. A test of the linearity of this relationship provided no evidence against linearity, F (39,32) = 1.0, p >.10. While subsequent analysis of the residuals showed some heterogeneity of variance, this was also eliminated by the transformations which removed skewness. However, since the correlational analysis is relatively unaffected by substitution of the transformed scores, for simplicity the reading time measure is left untransformed in the forthcoming analyses.

[4]We will use a standard format for these analyses -- estimated r (standard error), number of subjects upon which the analysis is based. The standard error pro-

vides a rough confidence interval for the estimate of r, plus or minus one standard deviation (i.e., 68%). Degrees of freedom for the χ^2 tests of fit are presented as a subscripted value.

[5]Several analyses depend on estimating the role of two-choice RT in the other RT measures. All of these measures are somewhat skewed, particularly two-choice RT (skew = .8), due to long outlying RTs. Repeating an earlier analysis used to test the distribution of the residuals from regression of these variables showed little sign of skewness. Similarly, little evidence of heterogeneity of variance or nonlinearity was found, so transformation was unnecessary.

APPENDIX A.

STIMULUS MATERIALS FOR THE INFORMATION PROCESSING TASKS

Four tables are included, which display the materials task by task. The materials in Tables A1, A2, and A4 originated in the present research; those in Table A3 were kindly provided by David Meyer (cf. Meyer et al., Note 8).

Table A1

Letter and Word Stimuli for the Visual Search, Span of Apprehension, and Memory Span Tasks

Letters	Words
a	lion
b	seat
c	lime
d	oven
e	iron
f	beer
g	tile
h	dish
j	wall
k	foil
l	barn
m	club
n	suit
o	base
p	coat
r	salt
s	sail
t	tent
u	ring
v	hand

Table A2

Letter and Word Stimuli for the Pairs Used in the Matching Task

Letters	Words			
a	sink	fake	wink	fare
b	date	dice	gate	dine
c	damp	gale	ramp	game
d	feed	rope	weed	role
e	buck	nose	tuck	note
j	cent	race	dent	rare
l	sore	mile	tore	mine
m	bake	wine	wake	wise
n	show	leaf	slow	leap
o	shot	harm	slot	harp
p	slat	care	spat	cart
r	shed	clap	sled	claw
s	shop	bark	slop	barn
t	slur	heal	spur	hear
w	swap	form	slap	fork
x	ship	wind	slip	wing

Letters	Words
w	home
x	bank
y	body
z	pipe

Table A3

The Word and Nonword Stimulus Pairs

for the Lexical Decision Task

Associates	Nonassociates
king - queen	uncle - silk
boy - man	train - heat
butter - bread	mind - cabbage
nail - hammar	meat - home
avenue - street	wool - city
baby - infant	apple - stream
lock - key	justice- hand
table - chair	seed - jar
peace - war	glass - dirt
cat - dog	oil - flower
earth - ground	church - doctor
sister - brother	grass - lamb
pin - needle	window - spool
salt - pepper	scissors-insect
carpet - rug	plant - bath
cup - saucer	waltz - bug

Word-Nonword	Nonword-Word	Nonword-Nonword
thread - yine	heg - yard	reater - blar
tiger - nong	fotsen - candy	dubter -voap

eye	- voon	firch	- adult	dake	- telt
moth	- zock	nool	- music	jind	- vake
bed	- rair	listract	- leaf	srow	- glain
moon	- shief	korse	- officer	rast	- selt
cotton	- bisic	enpet	- cup	plue	- flok
river	- vall	morkbage	- soldier	lail	- keat
lion	- fuass	yold	- mountain	glat	- vack
stove	- namb	shirser	- foot	zill	- woct
cracker	- erter	kipy	- health	daltz	- nitvy
soap	- pabing	jad	- knife	jeal	- tiew
thief	- huver	loitel	- plane	mitor	- stread
army	- wegaring	globble	- sickness	poy	- jasom
door	- strink	gow	- lettuce	speet	- oye
fruit	- tocan	katmy	- food	stroom	- dound
shoe	- rapger	poft	- hair	bither	- bance
web	- jilinue	ludge	- blossom	nour	- oit
star	- phalcle	fup	- dish	ded	- lome
law	- zat	crisow	- steak	fough	- ovigion
night	- joth	wimer	- cheese	lurse	- jarent
sea	- vut	udy	- wine	kog	- glem
dime	- cleep	hanwake	- whistle	ralt	- pleam
mouse	- ogtuce	eaple	- town	jeow	- glat
song	- strant	purch	- fox	veetain	- strapes
grapes	- goe	pheese	- spider	ceal	- krow
dance	- truit	tusier	- school	ralt	- lanyon

injury	- flove	knread	- water	selt	- prug
crown	- sutmy	fleep	- stem	dood	- phife
flag	- gneat	rinder	- pill	meb	- shair
animal	- breen	gloet	- toast	vigle	- geed
lamp	- eal	roise	- chalk	pirlow	- greal

Table A4

The Sentences Used in the Semantic Decision Task

True Sentences		
Level 0	Level 1	Level 2
A Chevrolet is a Chevrolet	A Corvette is a car	A Pontiac is a vehicle
Brandy is brandy	Whiskey is alcohol	Champagne is a beverage
A Bulldog is a Bulldog	A Collie is a dog	A Terrier is a pet
Beef is beef	Pork is meat	Veal is food
A Private is a Private	A Lieutenant is a soldier	A Corporal is a person
A Maple is a Maple	A Cedar is a tree	An Elm is a plant
Christmas is Christmas	Thanksgiving is a holiday	Easter is a day
A novel is a novel	A dictionary is a book	A Bible is reading material
An emerald is an emerald	A diamond is a jewel	A ruby is a stone
Football is football	Baseball is a sport	Golf is an activity
A cobra is a cobra	A copperhead is a snake	A python is a reptile
Coca-Cola is Coca-Cola	7Up is a soft drink	Sprite is a liquid
Halibut is halibut	Salmon is a fish	Smelt is a seafood
A gorilla is a gorilla	A chimpanzee is a primate	An orangutang is an ape

A canary is a canary	A sparrow is a bird	A robin is an animal
A father is a father	An uncle is a relative	A grandfather is a male
Princeton is Princeton	Oxford is a University	Yale is a school
Calculus is calculus	Trigonometry is math	Algebra is a school subject
A pistol is a pistol	A revolver is a gun	A rifle is a weapon
A mansion is a mansion	A cottage is a house	A bungalow is a building

False Sentences

Level 0	Level 1	Level 2
Toyota is a cardigan	Datsun is a sweater	A Ford is a clothing
A poodle is a Boston cream	An Irish setter is a pie	A Laborador is a dessert
Beer is a black widow	Vodka is a spider	Gin is an insect
Chicken is the Atlantic	Lamb is an ocean	Ham is water
A Sergeant is salt	A Captain is a spice	A Major is a seasoning
A Dogwood is a canoe	An alder is a boat	A fir is transportation
Memorial Day is cheddar	The 4th of July is a cheese	Labor Day is a milk product
A biography is Longfellow	A paperback is a poet	An encyclopedia is a writer
A sapphire is a minister	A pearl is a clergyman	An opal is a profession

Flounder is a clarinet	Tuna is a wind instrument	Cod is a musical instrument
Volleyball is a toe	Tennis is a part of a foot	Soccer is a part of a body
A cannon is a cantaloupe	A shotgun is a melon	An automatic is a fruit
Root Beer is a penny	Dr. Pepper is a coin	Orange Crush is money
Geometry is a loafer	Arithmetic is a shoe	Statistics is footwear
A blue jay is a knife	A crow is silverware	A finch is a utensil
A cabin is June	An adobe is a month	A hut is a unit of time
Harvard is butane	Cambridge is a gas	Stanford is a type of fuel
A brother is iron	A nephew is metal	A son is a mineral
A rattler is Mt. Rainier	A viper is a volcano	Boa is an earth formation
Monkey is aspirin	A Gibbon is a medicine	A Rhesus is a drug

APPENDIX B.

INTERCORRELATIONS OF PSYCHOMETRIC AND INFORMATION PROCESSING MEASURES

Three tables are included which summarize all of the relationships among the psychometric tests and the information processing measures. For convenience, Table B1 presents the correlations for just the 12 information processing measures, Table B2 displays the correlations for just the 16 psychometric measures, and Table B3 presents a 12 X 16 matrix of the intercorrelations among the psychometric and information processing measures. All are Pearson product-moment correlation coefficients based on the 67 subjects with complete data on all measures.

For Tables B1 and B2, reliability estimates appear in italics on the diagonal. Unless otherwise indicated, these represent split-half reliabilities based on odd versus even trials from our data. All correlation coefficients are positive for positive relationships to reduce confusion in their interpretation.

Table B1

Intercorrelations of the Information Processing Measures

TASK MEASURE		2CRT	MS-L	MS-W	SA-L	SA-W	VS-L	VS-W	M-1	M-W	LD	PSV	SD
Two-choice reaction time		.96/.80 [a]											
Memory Span	letters	.22	>.6[b]										
	words	.24	.63	>.6[b]									
Span of Apprehension	letters	.27	.03	.12	>.6[b]								
	words	.21	-.02	.20	.59	>.6[b]							
Visual Search	letters	.70	.25	.34	.26	.15	.94						
	words	.58	.08	.17	.26	.39	.67	.86					
Matching	letters	.68	.25	.34	.25	.24	.74	.62	.92				
	words	.65	.13	.20	.27	.45	.64	.81	.70	.94			
Lexical Decision		.63	.30	.34	.10	.20	.60	.62	.67	.71	.85[c]		
Picture-Sentence Verification		.51	.31	.35	.26	.23	.52	.57	.39	.51	.55	.94	
Semantic Decision		.62	.28	.34	.20	.40	.51	.67	.54	.76	.84	.64	.92

a The second value represents day-to-day reliability

b Lower bound based on claims in the test literature

c Underestimate based on correlation of first and second reaction times on a given trial

Table B2

Intercorrelations of the Psychometric Measures

Psychometric Measures	E-RT	ND-RT	M-RT	ND-RC	D-RC	D-LC	WPC-RC	WPC-Vocab	WPC-Gram	WPC-Spell	WPC-QS	WPC-QA	WPC-AQ	WPC-Mech	WPC-Spat	RM-GI
Experimental Reading Time	.92[a]															
Nelson-Denny Reading Time	.63	.77[b]														
Minnesota Reading Time	.50	.66	.79[c]													
Nelson-Denny Reading Comprehension	.60	.53	.37	>.8[d]												
Davis Reading Comprehension	.53	.40	.33	.72	.70[a]											
Davis Listening Comprehension	.52	.30	.29	.74	.80	.75[a]										
WPC Reading Comprehension	.46	.37	.25	.70	.64	.68	~.8[e]									
WPC Vocabulary	.49	.26	.30	.67	.68	.78	.67	~.8[e]								
WPC Grammar	.43	.19	.35	.53	.56	.62	.63	.59	~.8[e]							
WPC Spelling	.26	.13	.18	.26	.29	.24	.36	.38	.54	~.8[e]						
WPC Quantitative Skill	.29	.08	.11	.37	.36	.40	.39	.29	.46	.20	~.8[e]					
WPC Quantitative Achievement	.31	.13	.21	.28	.29	.40	.34	.31	.46	.33	.65	~.8[e]				
WPC Applied Quantitative	.42	.20	.30	.42	.40	.45	.40	.46	.41	.29	.78	.62	~.8[e]			
WPC Mechanical	.22	.09	.13	.35	.39	.56	.33	.53	.34	.06	.45	.39	.45	~.8[e]		
WPC Spatial	.08	.02	-.04	.11	.12	.22	.14	.19	.20	.14	.41	.46	.26	.47	~.8[e]	
Raven Matrices General Intelligence	.23	.14	.05	.33	.29	.39	.28	.23	.40	.14	.60	.56	.35	.45	.43	>.9[d]

[a] Underestimate based on combining parallel tests without correction for length.

[b] Based on parallel forms correlation in Hunt, McKee, and Yantis (1978).

[c] Reported in Holmes (1954).

[d] Lower bound based on claims in the test literature.

[e] According to the WPC Test Manual (Note 1).

Table B3

Intercorrelations of the Information Processing Measures and the Psychometric Measures

		Exp. RT	ND RT	MINN RT	ND RC	DAVIS RC	DAVIS LC	WPC RC	WPC VOCAB	WPC GRAM	WPC SPELL	WPC Q.SKILL	WPC Q.ACH	WPC AQ	WPC MECH	WPC SPAT	R GI
Two-Choice Reaction Time		.24	.17	.14	.23	.25	.21	.07	.07	.26	-.04	.16	.15	.19	.15	.07	.11
Memory Span	Letters	.08	.08	.14	.17	.15	.16	.26	.20	.31	.25	.09	.04	.05	.32	.20	.09
	Words	.15	.12	.10	.44	.35	.30	.39	.30	.42	.30	.20	.07	.10	.33	.11	.24
Span of Apprehension	Letters	.02	-.05	-.13	.05	-.04	.05	.11	.13	.10	.04	-.01	.09	.07	.20	.15	.15
	Words	.28	.17	.09	.31	.13	.22	.14	.28	.18	.18	.16	.28	.28	.21	.23	.17
Visual Search	Letters	.13	.06	.06	.07	.19	.09	-.05	-.10	.22	.05	.14	.24	.17	.19	.12	.28
	Words	.31	.18	.14	.20	.20	.14	.00	-.02	.20	.16	.14	.32	.27	.15	.19	.19
Matching	Letters	.14	.00	.02	.01	.03	-.04	.00	-.13	.16	-.05	.10	.15	.06	.08	.10	.22
	Words	.47	.30	.18	.29	.30	.19	.15	.16	.25	.29	.24	.35	.38	.16	.19	.17
Lexical Decision		.31	.27	.27	.30	.33	.18	.27	.16	.42	.30	.14	.27	.19	.08	.07	.20
Picture-Sentence Verification		.33	.24	.25	.36	.43	.39	.29	.28	.48	.34	.21	.35	.32	.34	.22	.32
Semantic Decision		.47	.39	.32	.47	.44	.35	.41	.39	.47	.45	.31	.45	.45	.28	.24	.24

TABLE 1

THE SCHEDULE OF TASKS

SESSION				
1	Two-Choice RT	Matching with letters	Visual Search with letters	
2	Two-Choice RT	Matching with words	Visual Search with words	
3	Two-Choice RT	Lexical Decision	Picture Sentence Verification	
4	Two-Choice RT	Semantic Decision	Inference	
5	Two-Choice RT	Span of Apprehension with letters and words	Memory Span with letters and words	Minnesota speed of reading test
6	Nelson-Denny reading rate and comprehension	Raven Matrices		
7	Modified-Davis comprehension tests: Listening	Experimental Reading Time Tests		
	Reading			
	Listening			
	Reading			

Table 2A

DESCRIPTIVE STATISTICS FOR THE REACTION-TIME MEASURES

MEASURE	OVERALL MEANS: MEDIAN RTs	MEAN RTs	STANDARD DEVIATIONS WITHIN-SUBJECTS	BETWEEN-SUBJECTS	PERCENT ERRORS	SPLIT-HALF RELIABILITY COEFFICIENTS
TWO-CHOICE REACTION TIME	271	273	42	29	2.0%	.96
SEARCH LETTERS	628	659	168	101	3.7%	.94
SEARCH WORDS	729	783	254	115	3.5%	.86
MATCHING LETTERS	568	590	128	69	4.1%	.92
MATCHING WORDS	815	835	170	108	4.5%	.94
LEXICAL DECISION	620	638	130	71	6.2%	.85
PICTURE-SENTENCE VERIFICATION	2022	2172	821	524	7.6%	.94
SEMANTIC DECISION	1043	1064	247	144	3.1%	.92

Note: The overall means are provided for median and mean reaction times of individual subjects. All measures are based upon 128 trials, with the exception of the Lexical Decision task (256) and Two-Choice Reaction Time (240). The reliability coefficient is the correlation between even and odd trials for all measures except Lexical Decision, where the correlation between the first and second block

TABLE 2B

DESCRIPTIVE STATISTICS FOR THE SPAN AND PSYCHOMETRIC MEASURES

MEASURE		MEANS	STANDARD DEVIATIONS BETWEEN-SUBJECTS	NUMBER OF ITEMS	RELIABILITY[a]
Memory Span	Letters	5.26	.80	15	>.6
	Words	5.14	.87	.5	>.6
Span of Apprehension	Letters	3.86	1.6	15	>.6
	Words	2.20	1.4	15	>.6
Experimental Reading Time		228	56	10	.92
Nelson-Denny Reading Time		217	84	1	.77
Minnesota Reading Time		353	103	36	.79
Davis Reading Comprehension		40.1	13.9	90	.70
Davis Listening Comprehension		45.5	15.8	90	.75
Nelson-Denny Reading Comprehension		22.9	6.0	36	.77
Raven Matrices Intelligence Test		24.7	4.8	36	.90

[a]More details on the sources of the reliability estimates appear in Appendix Tables A2

Table 3

READING TIME AND READING COMPREHENSION

CORRELATIONS WITH OTHER PSYCHOMETRIC MEASURES OF ABILITY

Measure	Reading Time	Reading Comprehension
Listening Comprehension	.52	.80
Vocabulary	.49	.68
Grammar	.43	.56
Spelling	.26	.29
Quantitative Skill	.29	.36
Quantitative Achievement	.31	.29
Applied Quantitative	.42	.40
Mechanical	.22	.39
Spatial	.08	.12
Raven	.23	.29

Table 4

READING TIME AND READING COMPREHENSION CORRELATIONS WITH THE VISUAL INFORMATION PROCESSING TASKS (SEPARATELY FOR LETTERS AND WORDS)

TASKS		READING TIME	READING COMPREHENSION
LETTERS	Span	.02	-.04
	Search	.13	.19
	Match	.14	.03
WORDS	Span	.31	.20
	Search	.47	.30
	Match	.31	.33

Table 5

INTERCORRELATIONS BETWEEN LETTER AND WORD VERSIONS OF THE VISUAL INFORMATION PROCESSING TASKS

		LETTERS		
		SPAN	SEARCH	MATCH
WORDS	SPAN	.59	.26	.27
	SEARCH	.15	.67	.64
	MATCH	.24	.62	.70

Table 6

READING TIME AND READING COMPREHENSION CORRELATIONS WITH THE HIGHER-ORDER INFORMATION PROCESSING TASKS

TASK	READING TIME	READING COMPREHENSION
Lexical Decision	.31	.33
Picture-sentence Verification	.33	.43
Semantic Decision	.47	.44

Table 7

Correlations Among Composite Variables Representing The Letter, Word, and Higher-Order Tasks, Together with Two-Choice Reaction Time

COMPOSITE VARIABLES	LETTERS	WORDS	HIGHER ORDER	TWO CHOICE RT
LETTERS	*.97*			
WORDS	.70	*.95*		
HIGHER-ORDER	.60	.77	*.96*	
TWO-CHOICE RT	.64	.53	.61	*.97*

Note: Composite variables are based on the sums of standardized scores for three tasks in each case (except two-choice RT). Reliabilities are shown in italics on the diagonal. The analyses are based on 86 subjects.

FIGURE CAPTIONS

Figure 1. Three applications of Joreskog's structural analysis of covariance matrices. Panel A shows a simple, one-trait model; Panel B shows the general case of a two-trait model. Panel C is a specific three-trait model appropriate for our situation.

Figure 2. The distribution of Nelson-Denny reading comprehension scores for the 67 subjects included in our sample. For comparison purposes, these are plotted as a function of the appropriate normative data -- percentile ranks for the 14th grade.

Figure 3. Means of median reaction times and percentage of errors for the visual search task as a function of search set size. The curve parameter represents whether the target was or was not in the presented search set. Darkened bars represent in-set errors. The data are presented separately for letters and words.

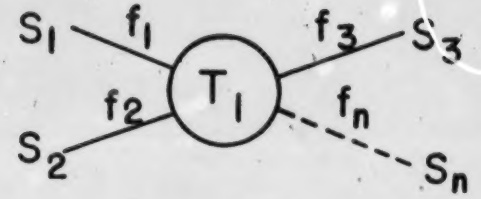
Figure 4. Means of median reaction times and percentage of errors for the picture-sentence verification task as a function of sentence-picture relationship (expresssed in terms of number of constituent comparisons in the Carpenter and Just model). The data are displayed separately for trials

where the first term in the sentence is identical to the top element in the picture (MATCH), and for trials where the first term in the sentence is identical to the bottom element in the picture (MISMATCH).

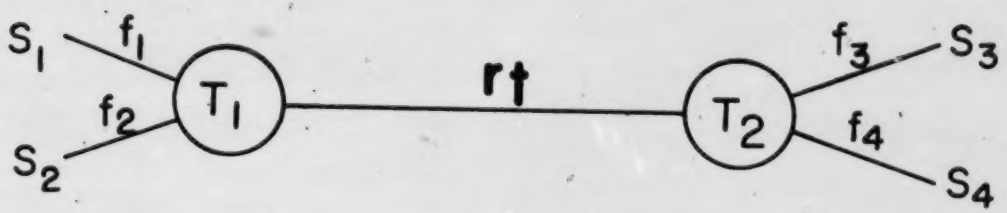
Figure 5. Means of median reaction times and percentage of errors for the semantic decision task as a function of level of hierarchical inclusion. The data are shown separately for true and false trials.

Figure 6. A scatter plot of the Davis listening comprehension scores against the Davis reading comprehension scores for our sample of 67 subjects, demonstrating their strong relationship.

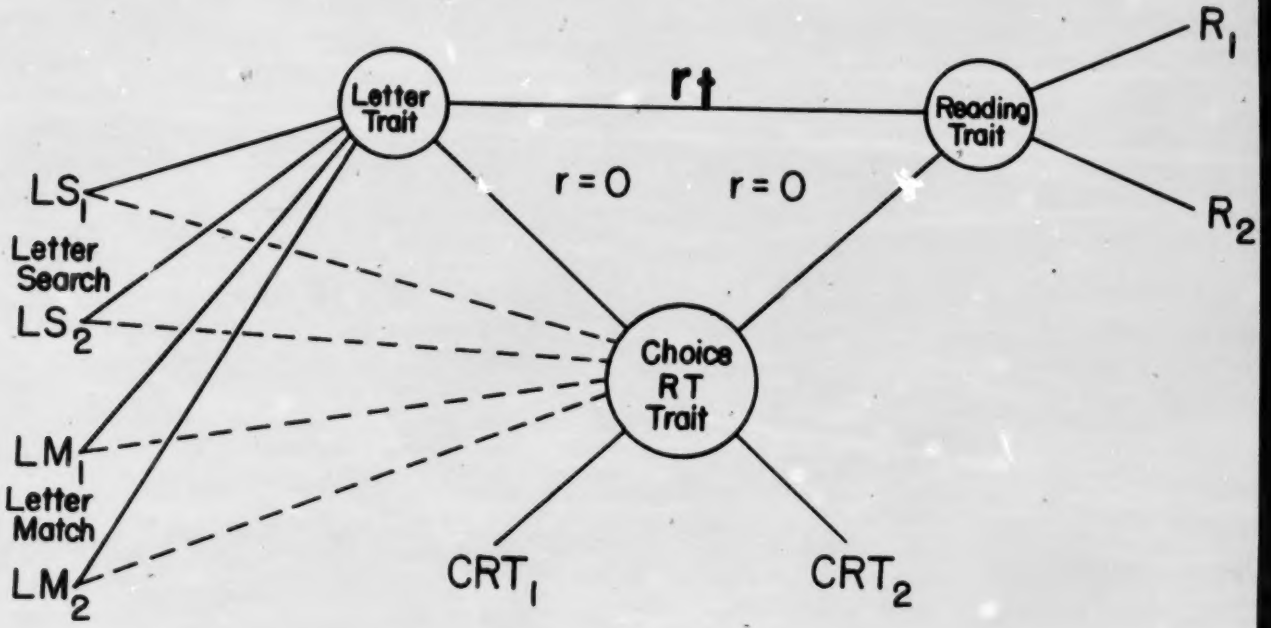
(A)

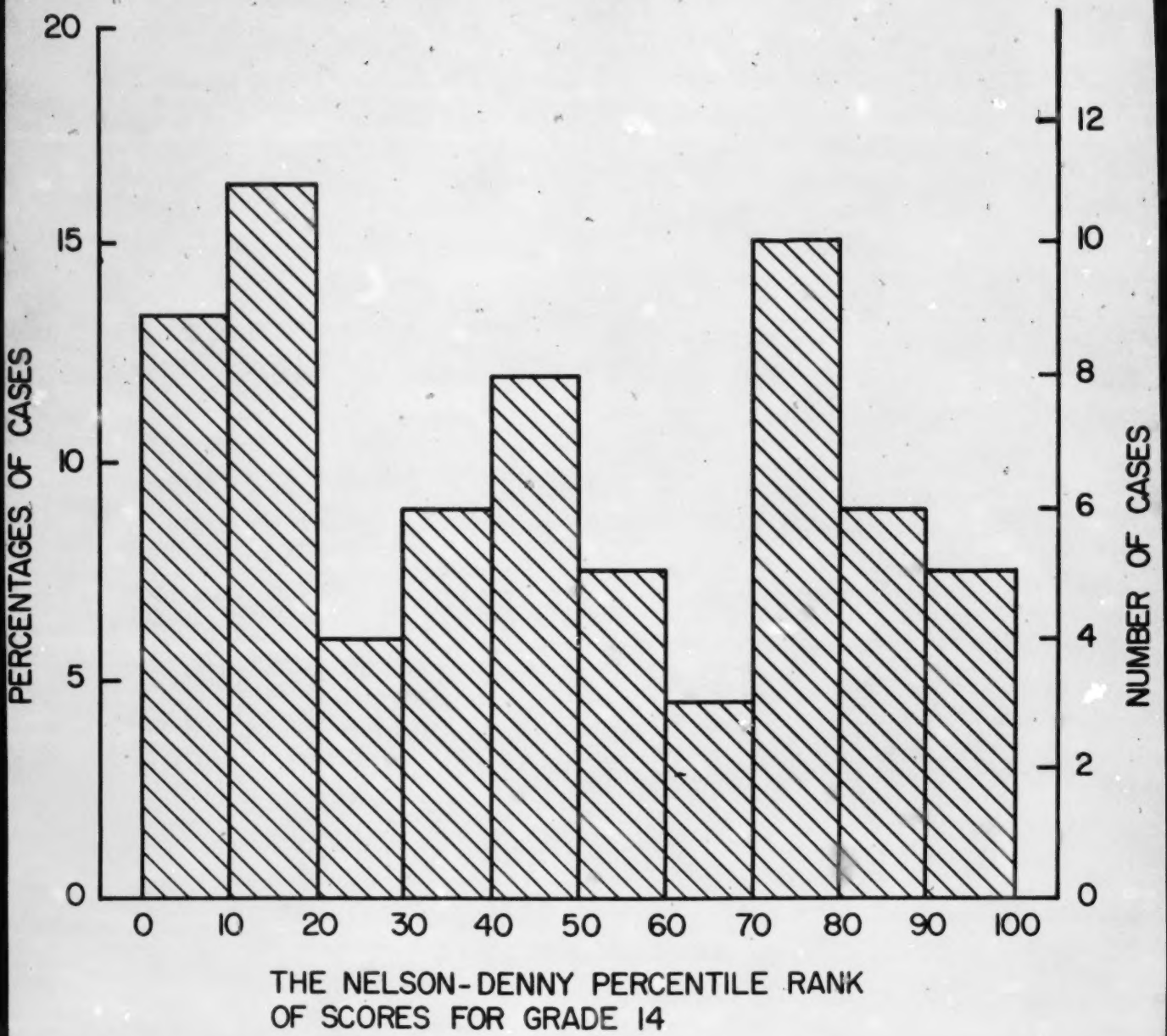

S1
f1
f3
S3
T1
f2
fn
S2
Sn

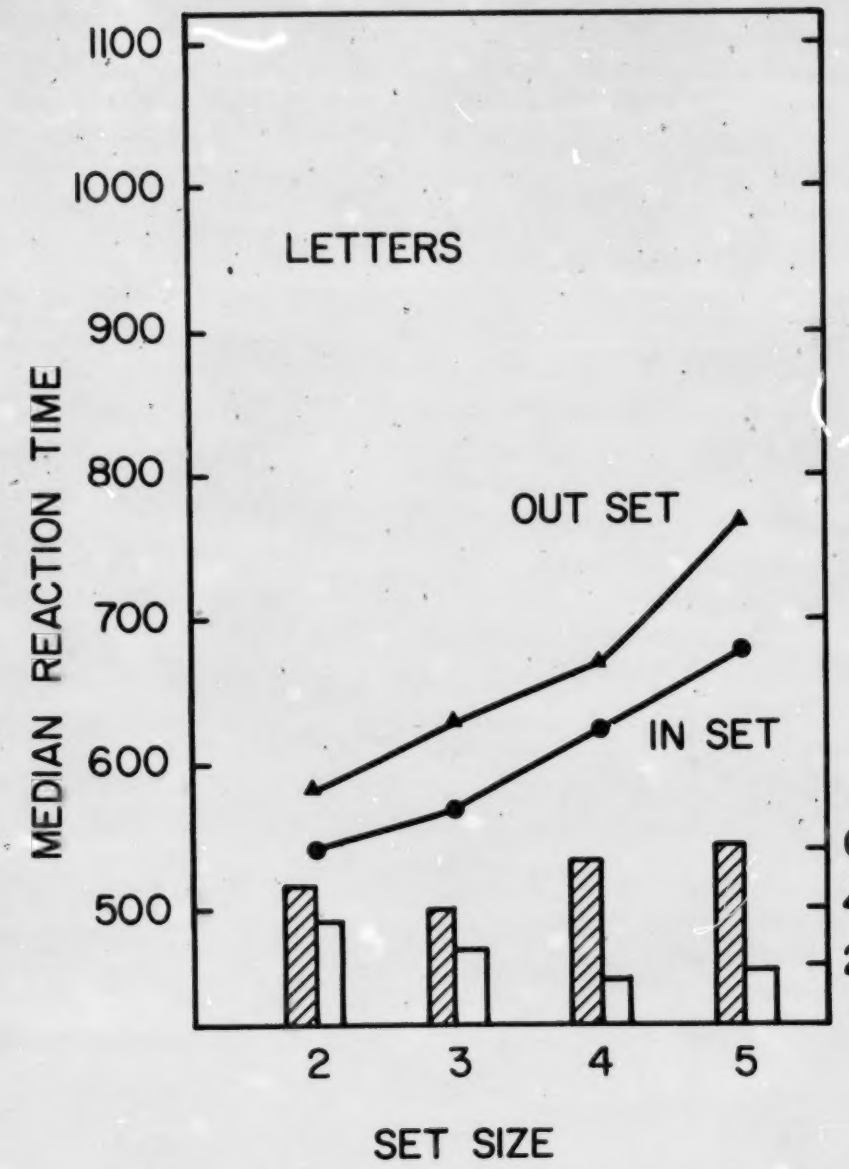
(B)


S1
f1
T1
rt
T2
f3
S3
f2
f4
S2
S4

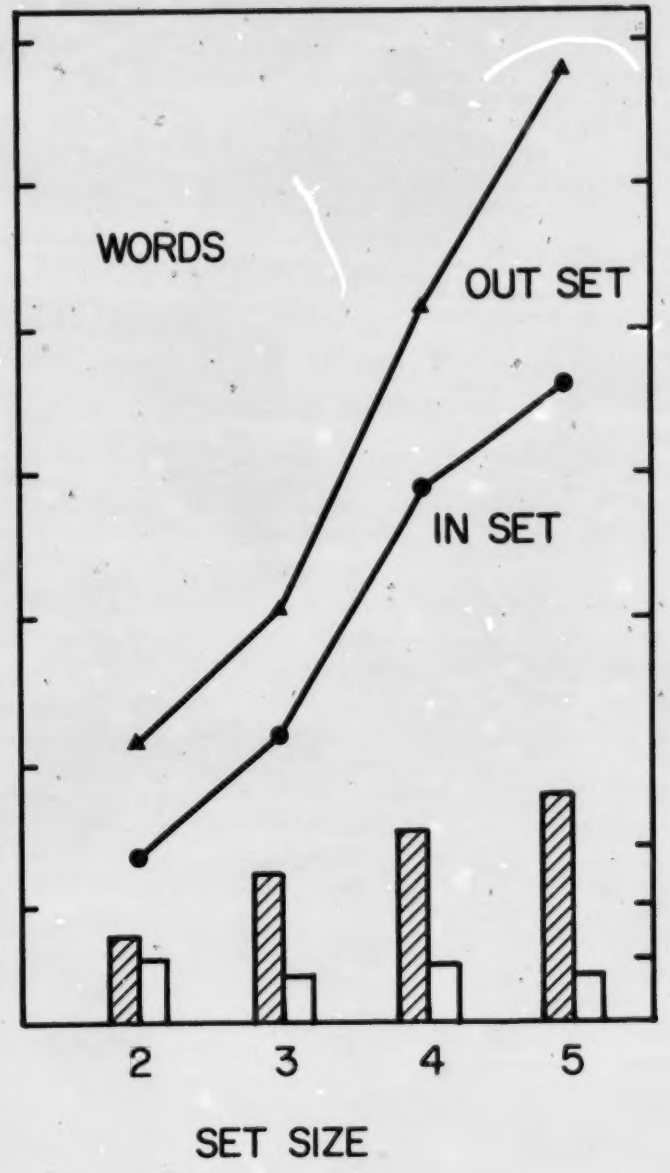
(C)


Letter Trait
rt
Reading Trait
R1
R2
r = 0
r = 0
LS1
Letter Search
LS2
Choice RT Trait
LM1
Letter Match
LM2
CRT1
CRT2

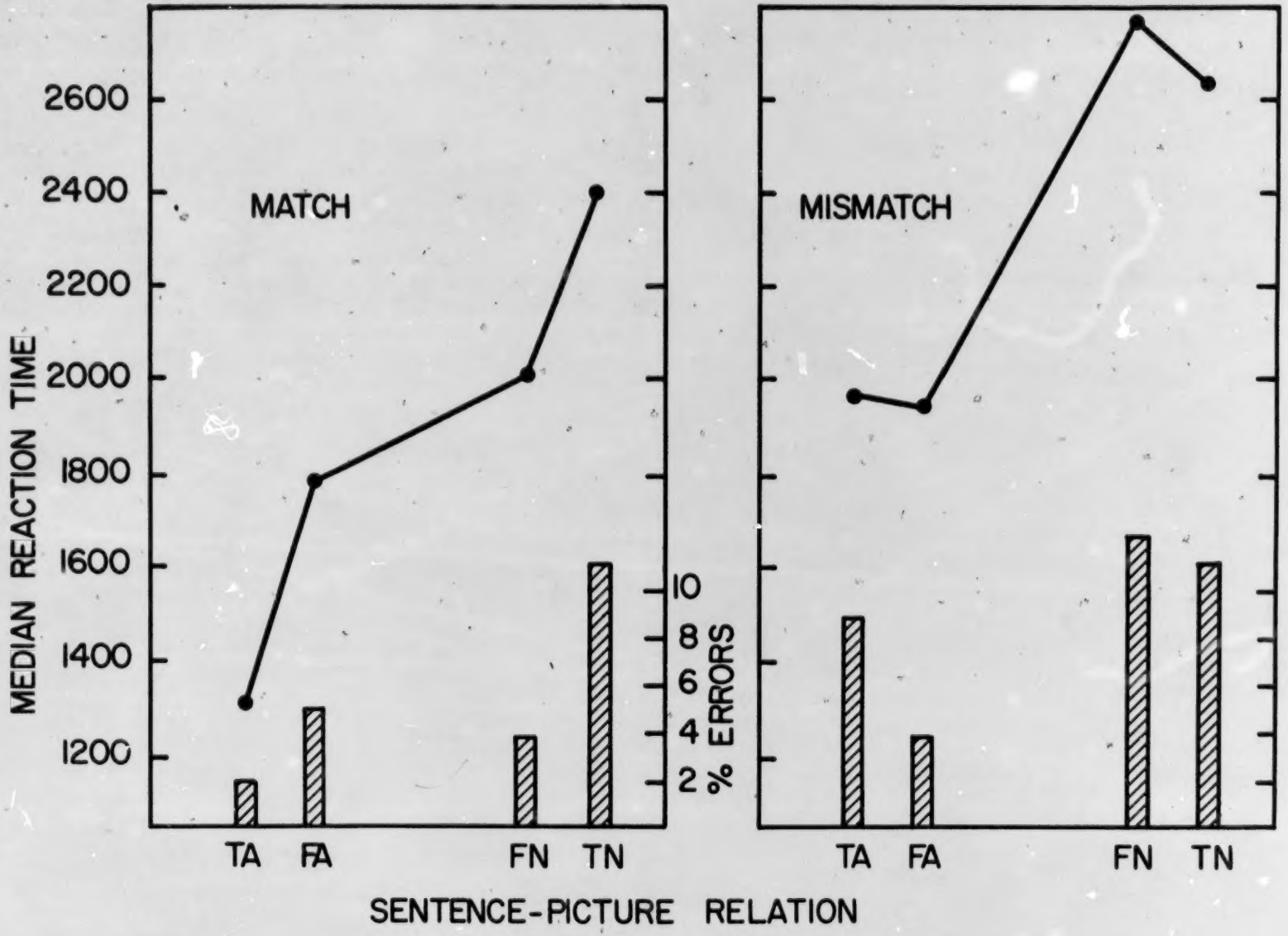

20
15
10
5
0
PERCENTAGES OF CASES
12
10
8
6
4
2
0
NUMBER OF CASES
0
10
20
30
40
50
60
70
80
90
100
THE NELSON-DENNY PERCENTILE RANK
OF SCORES FOR GRADE 14

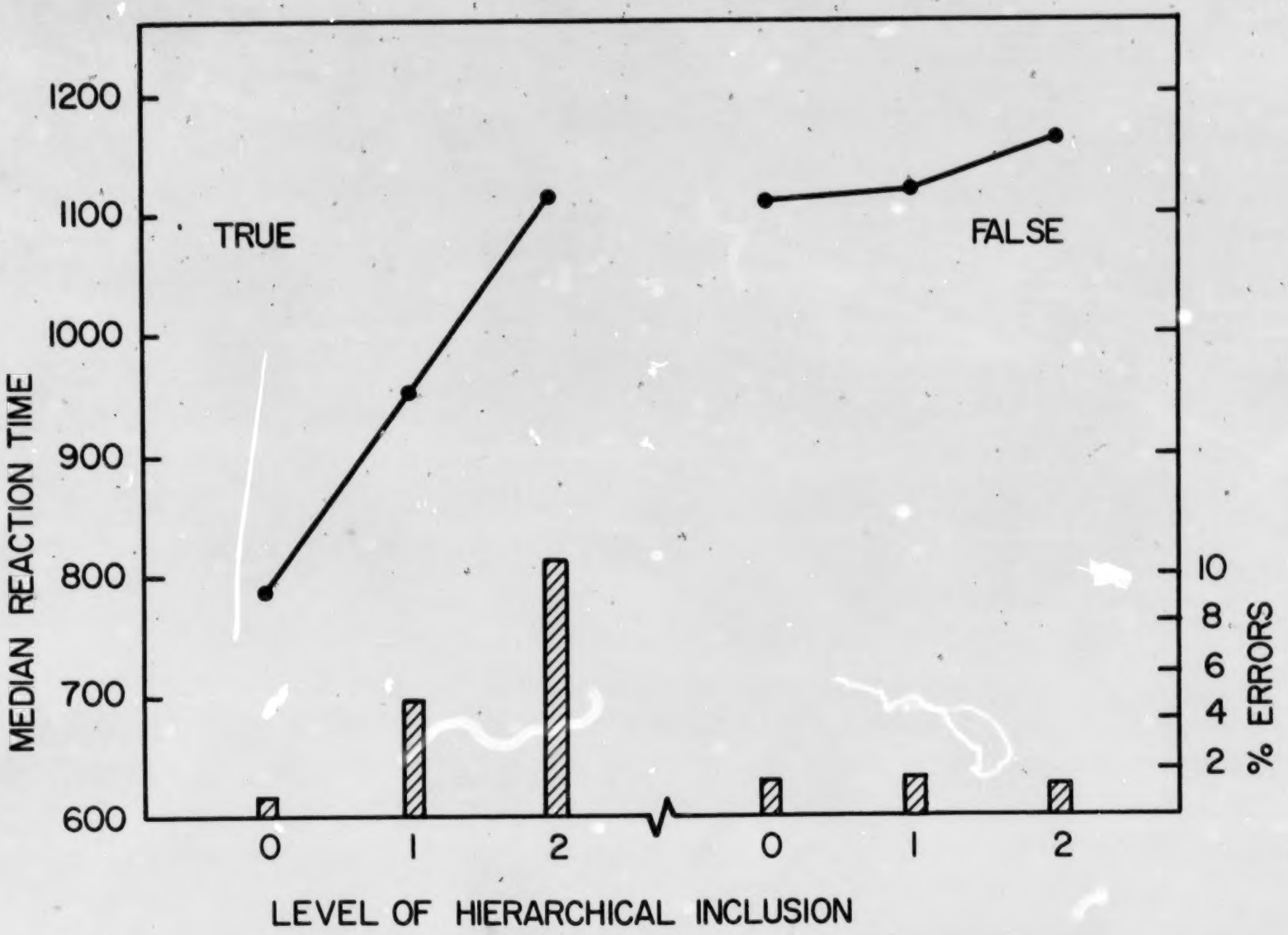

1100
1000
900
800
700
600
500
MEDIAN REACTION TIME
LETTERS
OUT SET
IN SET
6
4
2
% ERRORS
2
3
4
5
SET SIZE

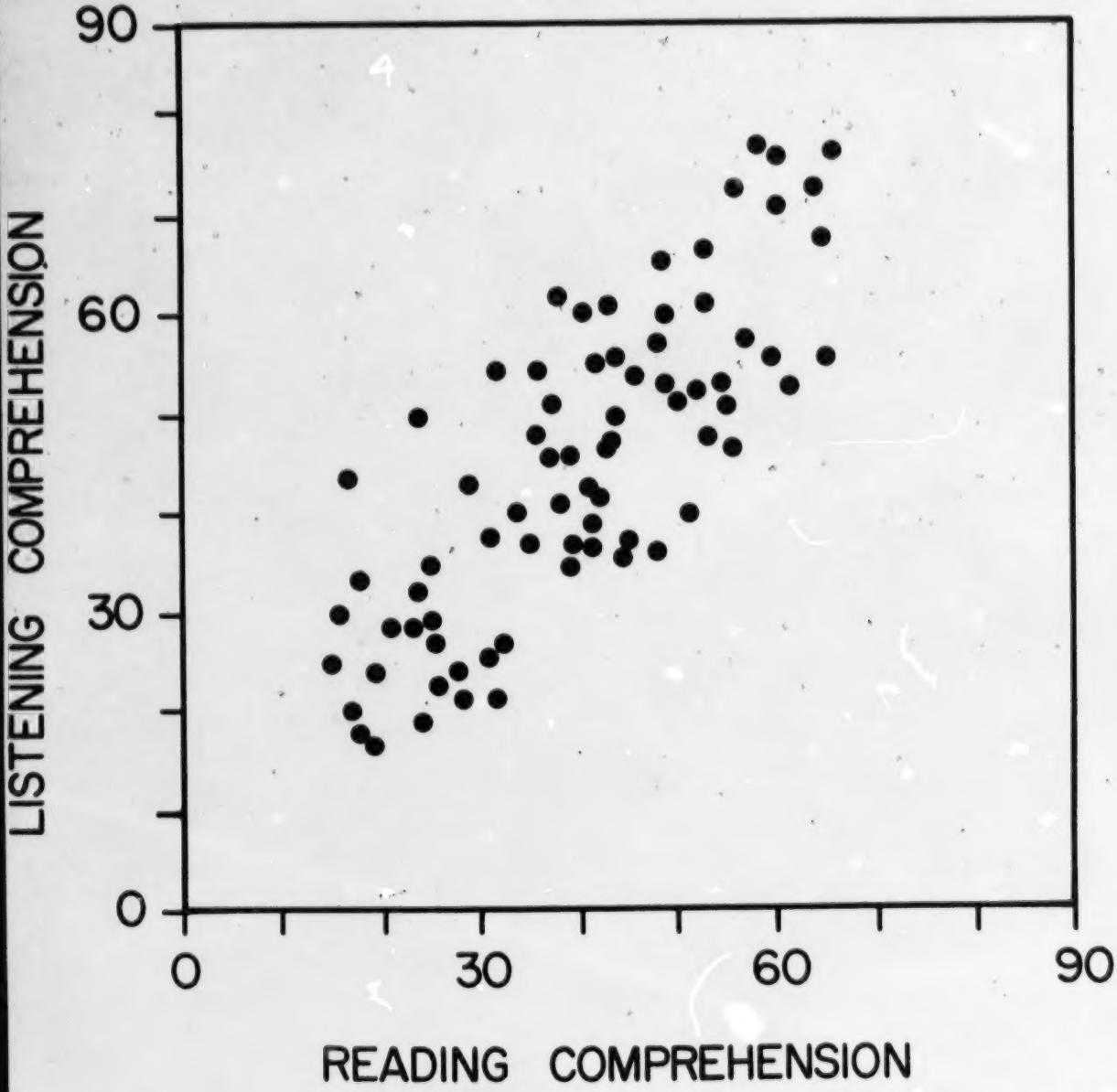
126


WORDS
OUT SET
IN SET
2
3
4
5
SET SIZE

127


MATCH
MISMATCH
MEDIAN REACTION TIME
2600
2400
2200
2000
1800
1600
1400
1200
% ERRORS
10
8
6
4
2
TA
FA
FN
TN
SENTENCE-PICTURE RELATION


1200
1100
1000
900
800
700
600
MEDIAN REACTION TIME
TRUE
FALSE
10
8
6
4
2
% ERRORS
0
1
2
0
1
2
LEVEL OF HIERARCHICAL INCLUSION


90
60
30
0
LISTENING COMPREHENSION
0
30
60
90
READING COMPREHENSION